GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 10, 2006, 09:32:19 ; Search time 163 Seconds
(without alignments)
1432.699 Million cell updates/sec

Title: US-10-646-470-1
Perfect score: 1806
Sequence: 1 MKRLVCVLLVCSSAVAQLHK.....MARNKGNHCGIASFPSYPEI 331

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_05.80:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1800	99.7	331	2	Q5T5I0_HUMAN	Q5t5i0 homo sapien
2	1797	99.5	331	1	CATS_HUMAN	P25774 homo sapien
3	1792	99.2	331	2	Q6FHS5_HUMAN	Q6fhs5 homo sapien
4	1700.5	94.2	330	1	CATS_SAIBB	Q8hy82 saimiri bol
5	1599	88.5	331	1	CATS_CANFA	Q8hy81 canis famil
6	1395.5	77.3	341	2	Q6PCU5_RAT	Q6pcu5 rattus norv
7	1363.5	75.5	340	2	Q99M14_MOUSE	Q99m14 mus musculu
8	1363.5	75.5	342	2	Q8BSZ5_MOUSE	Q8bsz5 mus musculu
9	1352.5	74.9	340	1	CATS_MOUSE	O70370 mus musculu
10	1299.5	72.0	330	1	CATS_RAT	Q02765 rattus norv
11	1216.5	67.4	328	2	Q5ZMR6_CHICK	Q5zmr6 gallus gall
12	1147	63.5	333	2	Q6DJC1_XENTR	Q6djc1 xenopus tro
13	1124.5	62.3	320	2	Q6DE57_XENLA	Q6de57 xenopus lae
14	1109.5	61.4	333	2	Q7T0S4_XENLA	Q7t0s4 xenopus lae
15	1086	60.1	334	2	Q7T183_9CICH	Q7t183 paralabidoc
16	1059	58.6	330	2	Q502A6_BRARE	Q502a6 brachydanio
17	1054	58.4	330	2	Q502H6_BRARE	Q502h6 brachydanio
18	1036	57.4	330	2	Q566T8_BRARE	Q566t8 brachydanio
19	1035	57.3	217	1	CATS_BOVIN	P25326 bos taurus
20	1033	57.2	330	2	Q4QRH8_BRARE	Q4qrh8 brachydanio
21	1029.5	57.0	337	2	Q58HF5_PAROL	Q58hf5 paralichthy
22	1025	56.8	337	2	Q6JZZ4_FUNHE	Q6jzz4 fundulus he
23	1017	56.3	331	2	Q90324_CYPCA	Q90324 cyprinus ca
24	1001	55.4	334	2	Q4SW28_TETNG	Q4sw28 tetraodon n
25	975.5	54.0	327	2	Q6F6A0_ORYLA	Q6f6a0 oryzias lat
26	959.5	53.1	330	1	CATK_PIG	Q9gle3 sus scrofa
27	955.5	52.9	334	2	Q5E968_BOVIN	Q5e968 bos taurus
28	952.5	52.7	329	1	CATK_RABIT	P43236 oryctolagus
29	951.5	52.7	329	1	CATK_HUMAN	P43235 homo sapien
30	951.5	52.7	329	1	CATK_MACFA	P61276 macaca fasc
31	951.5	52.7	329	1	CATK_MACMU	P61277 macaca mula
32	951.5	52.7	329	2	Q6FHN2_HUMAN	Q6fhn2 homo sapien
33	951.5	52.7	329	2	Q6FHS6_HUMAN	Q6fhs6 homo sapien
34	951	52.7	329	1	CATK_RAT	O35186 rattus norv
35	935	51.8	333	2	Q6PA77_XENLA	Q6pa77 xenopus lae
36	933	51.7	329	2	Q6DJA7_XENTR	Q6dja7 xenopus tro
37	931	51.6	329	1	CATK_MOUSE	P55097 mus musculu
38	931	51.6	329	2	Q545T0_MOUSE	Q545t0 mus musculu
39	931	51.6	355	2	Q5BL90_XENTR	Q5bl90 xenopus tro
40	928.5	51.4	337	2	Q6XRF7_FUNHE	Q6xrf7 fundulus he
41	922	51.1	336	2	Q6F6A1_ORYLA	Q6f6a1 oryzias lat
42	918	50.8	337	2	Q75SZ8_CYPCA	Q75sz8 cyprinus ca
43	914	50.6	338	2	Q90WC2_ONCMY	Q90wc2 oncorhynchu
44	910.5	50.4	335	2	Q68F41_XENLA	Q68f41 xenopus lae
45	906	50.2	336	2	Q4SW27_TETNG	Q4sw27 tetraodon n

ALIGNMENTS

RESULT 1
Q5T5I0_HUMAN
ID Q5T5I0_HUMAN PRELIMINARY; PRT; 331 AA.
AC Q5T5I0;
DT 01-FEB-2005 (TrEMBLrel. 29, Created)
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Cathepsin S.
GN Name=CTSS; ORFNames=RP11-363I22.2-001;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Howden P.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AL356292; CAI13657.1; -; Genomic_DNA.
DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
SQ SEQUENCE 331 AA; 37495 MW; 86093619DB6F0269 CRC64;

Query Match 99.7%; Score 1800; DB 2; Length 331;
Best Local Similarity 99.7%; Pred. No. 1.4e-141;
Matches 330; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              |||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
```

RESULT 2
CATS_HUMAN

```
ID   CATS_HUMAN     STANDARD;      PRT;   331 AA.
AC   P25774; Q9BUG3;
DT   01-MAY-1992 (Rel. 22, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Cathepsin S precursor (EC 3.4.22.27).
GN   Name=CTSS;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Alveolar macrophage;
RX   MEDLINE=92218373; PubMed=1373132;
RA   Shi G.-P., Munger J.S., Meara J.P., Rich D.H., Chapman H.A.;
RT   "Molecular cloning and expression of human alveolar macrophage
RT   cathepsin S, an elastinolytic cysteine protease.";
RL   J. Biol. Chem. 267:7258-7262(1992).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=94209337; PubMed=8157683;
RA   Shi G.-P., Webb A.C., Foster K.E., Knoll J.H.M., Lemere C.A.,
RA   Munger J.S., Chapman H.A.;
RT   "Human cathepsin S: chromosomal localization, gene structure, and
RT   tissue distribution.";
RL   J. Biol. Chem. 269:11530-11536(1994).
RN   [3]
RP   NUCLEOTIDE SEQUENCE [MRNA].
RC   TISSUE=Testis;
RX   MEDLINE=92317106; PubMed=1377692;
RA   Wiederanders B., Broemme D., Kirschke H., von Figura K., Schmidt B.,
RA   Peters C.;
RT   "Phylogenetic conservation of cysteine proteinases. Cloning and
RT   expression of a cDNA coding for human cathepsin S.";
RL   J. Biol. Chem. 267:13708-13713(1992).
RN   [4]
RP   SEQUENCE REVISION TO 211.
RA   Wiederanders B., Broemme D., Kirschke H., von Figura K., Schmidt B.,
RA   Peters C.;
RL   Submitted (MAY-1995) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC   TISSUE=Pancreas;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [6]
RP   3D-STRUCTURE MODELING OF 115-331.
RX   MEDLINE=99092748; PubMed=9876921; DOI=10.1093/protein/11.11.1007;
RA   Fengler A., Brandt W.;
RT   "Three-dimensional structures of the cysteine proteases cathepsins K
RT   and S deduced by knowledge-based modelling and active site
RT   characteristics.";
RL   Protein Eng. 11:1007-1013(1998).
CC   -!- FUNCTION: Thiol protease. Key protease responsible for the removal
CC       of the invariant chain from MHC class II molecules. The bond-
CC       specificity of this proteinase is in part similar to the
CC       specificities of cathepsin L and cathepsin N.
CC   -!- CATALYTIC ACTIVITY: Similar to cathepsin L, but with much less
CC       activity on Z-Phe-Arg-|-NHMec, and more activity on the Z-Val-Val-
CC       Arg-|-Xaa compound.
CC   -!- SUBUNIT: Monomer.
CC   -!- SUBCELLULAR LOCATION: Lysosomal.
CC   -!- SIMILARITY: Belongs to the peptidase C1 family.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; S93414; AAB22005.1; -; mRNA.
DR   EMBL; M86553; AAA35655.1; -; mRNA.
DR   EMBL; U07374; AAB60643.2; -; Genomic_DNA.
DR   EMBL; U07370; AAB60643.2; JOINED; Genomic_DNA.
DR   EMBL; U07371; AAB60643.2; JOINED; Genomic_DNA.
DR   EMBL; U07372; AAB60643.2; JOINED; Genomic_DNA.
DR   EMBL; U07373; AAB60643.2; JOINED; Genomic_DNA.
DR   EMBL; M90696; AAC37592.1; -; mRNA.
DR   EMBL; BC002642; AAH02642.1; -; mRNA.
DR   PIR; A42482; A42482.
DR   PDB; 1BXF; Model; A=115-331.
DR   PDB; 1GLO; X-ray; A=115-331.
DR   PDB; 1MS6; X-ray; A=115-331.
DR   PDB; 1NPZ; X-ray; A/B=115-331.
DR   PDB; 1NQC; X-ray; A=115-331.
DR   MEROPS; C01.034; -.
DR   Ensembl; ENSG00000163131; Homo sapiens.
DR   HGNC; HGNC:2545; CTSS.
DR   H-InvDB; HIX0001035; -.
DR   MIM; 116845; -.
DR   GO; GO:0005576; C:extracellular region; NAS.
DR   GO; GO:0005764; C:lysosome; NAS.
DR   GO; GO:0004218; F:cathepsin S activity; TAS.
DR   GO; GO:0006955; P:immune response; TAS.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; TAS.
DR   InterPro; IPR000169; Pept_cys_AS.
DR   InterPro; IPR000668; Peptidase_C1.
DR   PANTHER; PTHR12411; Peptidase_C1; 1.
DR   Pfam; PF00112; Peptidase_C1; 1.
DR   PRINTS; PR00705; PAPAIN.
DR   ProDom; PD000158; Peptidase_C1; 1.
DR   SMART; SM00645; Pept_C1; 1.
DR   PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
DR   PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
KW   3D-structure; Glycoprotein; Hydrolase; Lysosome; Protease; Signal;
KW   Thiol protease; Zymogen.
FT   SIGNAL       1    16       Potential.
FT   PROPEP      17   114       Activation peptide.
FT   CHAIN      115   331       Cathepsin S.
FT   ACT_SITE   139   139       By similarity.
FT   ACT_SITE   278   278       By similarity.
FT   ACT_SITE   298   298       By similarity.
FT   CARBOHYD   104   104       N-linked (GlcNAc. . .) (Potential).
FT   DISULFID   126   224
FT   DISULFID   136   180
FT   DISULFID   170   213
FT   DISULFID   272   320
FT   CONFLICT    92    92       M -> T (in Ref. 1 and 2).
FT   CONFLICT   113   113       R -> W (in Ref. 5).
FT   CONFLICT   161   161       T -> S (in Ref. 3 and 5).
FT   STRAND     119   120
FT   HELIX      121   124
FT   TURN       125   125
FT   STRAND     132   132
```

```
FT   TURN        134    135
FT   STRAND      137    137
FT   HELIX       139    156
FT   STRAND      162    162
FT   HELIX       164    170
FT   HELIX       173    175
FT   TURN        176    176
FT   TURN        179    180
FT   STRAND      181    181
FT   HELIX       185    195
FT   TURN        196    196
FT   STRAND      198    200
FT   TURN        201    203
FT   HELIX       217    219
FT   STRAND      220    222
FT   STRAND      226    229
FT   TURN        232    233
FT   HELIX       235    244
FT   TURN        245    245
FT   STRAND      248    252
FT   HELIX       257    261
FT   STRAND      266    267
FT   TURN        270    271
FT   STRAND      278    288
FT   TURN        289    290
FT   STRAND      291    297
FT   STRAND      300    300
FT   TURN        302    303
FT   TURN        305    305
FT   STRAND      306    306
FT   TURN        307    308
FT   STRAND      309    313
FT   HELIX       319    321
FT   TURN        322    324
FT   STRAND      327    331
SQ   SEQUENCE   331 AA;  37510 MW;  835935FA56B78902 CRC64;

  Query Match             99.5%;  Score 1797;  DB 1;  Length 331;
  Best Local Similarity   99.4%;  Pred. No. 2.5e-141;
  Matches  329;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||:|||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVTLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              |||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331


RESULT 3
Q6FHS5_HUMAN
ID   Q6FHS5_HUMAN PRELIMINARY;      PRT;   331 AA.
AC   Q6FHS5;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   CTSS protein.
GN   Name=CTSS;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Halleck A., Ebert L., Mkoundinya M., Schick M., Eisenstein S.,
RA   Neubert P., Kstrang K., Schatten R., Shen B., Henze S., Mar W.,
RA   Korn B., Zuo D., Hu Y., LaBaer J.;
RL   Submitted (JUN-2004) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Belongs to the peptidase C1 family.
DR   EMBL; CR541676; CAG46477.1; -; mRNA.
DR   SMR; Q6FHS5; 115-331.
DR   GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000668; Peptidase_C1.
DR   InterPro; IPR000169; Pept_cys_AS.
DR   Pfam; PF00112; Peptidase_C1; 1.
DR   PRINTS; PR00705; PAPAIN.
DR   ProDom; PD000158; Peptidase_C1; 1.
DR   SMART; SM00645; Pept_C1; 1.
DR   PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
DR   PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
KW   Hydrolase; Protease; Thiol protease.
SQ   SEQUENCE   331 AA;  37526 MW;  814F433BDC1A223B CRC64;

  Query Match             99.2%;  Score 1792;  DB 2;  Length 331;
  Best Local Similarity   99.4%;  Pred. No. 6.6e-141;
  Matches  329;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNWILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              |||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331


RESULT 4
CATS_SAIBB
ID   CATS_SAIBB     STANDARD;      PRT;   330 AA.
AC   Q8HY82;
DT   10-MAY-2005 (Rel. 47, Created)
DT   10-MAY-2005 (Rel. 47, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Cathepsin S precursor (EC 3.4.22.27).
GN   Name=CTSS;
OS   Saimiri boliviensis boliviensis (Bolivian squirrel monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Platyrrhini; Cebidae;
OC   Cebinae; Saimiri.
OX   NCBI_TaxID=39432;
```

```
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=22538941; PubMed=12651112; DOI=10.1016/S1046-5928(02)00646-0;
RA   Baker S.M., Karlsson L., Thurmond R.L.;
RT   "Cloning, expression, purification, and activity of dog (Canis
RT   familiaris) and monkey (Saimiri boliviensis) cathepsin S.";
RL   Protein Expr. Purif. 28:93-101(2003).
CC   -!- FUNCTION: Thiol protease. Key protease responsible for the removal
CC       of the invariant chain from MHC class II molecules. The bond-
CC       specificity of this proteinase is in part similar to the
CC       specificities of cathepsin L and cathepsin N (By similarity).
CC   -!- CATALYTIC ACTIVITY: Similar to cathepsin L, but with much less
CC       activity on Z-Phe-Arg-|-NHMec, and more activity on the Z-Val-Val-
CC       Arg-|-Xaa compound.
CC   -!- SUBCELLULAR LOCATION: Lysosomal (By similarity).
CC   -!- SIMILARITY: Belongs to the peptidase C1 family.
CC   --------------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   --------------------------------------------------------------------------
DR   EMBL; AY156691; AAO13008.1; -; mRNA.
DR   HSSP; P25774; 1MS6.
DR   SMR; Q8HY82; 115-330.
DR   MEROPS; C01.034; -.
DR   InterPro; IPR000169; Pept_cys_AS.
DR   InterPro; IPR000668; Peptidase_C1.
DR   PANTHER; PTHR12411; Peptidase_C1; 1.
DR   Pfam; PF00112; Peptidase_C1; 1.
DR   PRINTS; PR00705; PAPAIN.
DR   ProDom; PD000158; Peptidase_C1; 1.
DR   SMART; SM00645; Pept_C1; 1.
DR   PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
DR   PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
KW   Glycoprotein; Hydrolase; Lysosome; Protease; Signal; Thiol protease;
KW   Zymogen.
FT   SIGNAL        1     17       Potential.
FT   PROPEP       18    114       Activation peptide (By similarity).
FT   CHAIN       115    330       Cathepsin S.
FT   ACT_SITE    139    139       By similarity.
FT   ACT_SITE    277    277       By similarity.
FT   ACT_SITE    297    297       By similarity.
FT   CARBOHYD    104    104       N-linked (GlcNAc. . .) (Potential).
FT   DISULFID    126    223       By similarity.
FT   DISULFID    136    179       By similarity.
FT   DISULFID    170    212       By similarity.
FT   DISULFID    271    319       By similarity.
SQ   SEQUENCE   330 AA;  37346 MW;  4ECE3129F2418C34 CRC64;

  Query Match             94.2%;  Score 1700.5;  DB 1;  Length 330;
  Best Local Similarity   93.7%;  Pred. No. 2.8e-133;
  Matches  310;  Conservative   11;  Mismatches    9;  Indels    1;  Gaps    1;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||:||||| |||||| ||||||||||||:|||||||||||||||||||||||||||||||
Db          1 MKQLVCVLFVCSSAVTQLHKDPTLDHHWNLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              ||||||||||||||||||||||||||||||||||||||:|||||||||||||::||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPNQWQRNITYKSNPNQMLPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCS-EKYGNKGC 179

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              |||||| |||||||||||||:||||||| | |||||||||||||||||||||||||||||
Db        180 NGGFMTEAFQYIIDNKGIDSEASYPYKATDQKCQYDSKYRAATCSKYTELPYGREDVLKE 239

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              |||||||| ||||| |||||||||||||:|:||| |||||||:|||||||||||||||||
Db        240 AVANKGPVCVGVDASHPSFFLYRSGVYYDPACTQKVNHGVLVIGYGDLNGKEYWLVKNSW 299

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              | ||||:|||||||||||||||||:||||||
Db        300 GSNFGEQGYIRMARNKGNHCGIASYPSYPEI 330


RESULT 5
CATS_CANFA
ID   CATS_CANFA     STANDARD;      PRT;   331 AA.
AC   Q8HY81;
DT   10-MAY-2005 (Rel. 47, Created)
DT   10-MAY-2005 (Rel. 47, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Cathepsin S precursor (EC 3.4.22.27).
GN   Name=CTSS;
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
OC   Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=22538941; PubMed=12651112; DOI=10.1016/S1046-5928(02)00646-0;
RA   Baker S.M., Karlsson L., Thurmond R.L.;
RT   "Cloning, expression, purification, and activity of dog (Canis
RT   familiaris) and monkey (Saimiri boliviensis) cathepsin S.";
RL   Protein Expr. Purif. 28:93-101(2003).
CC   -!- FUNCTION: Thiol protease. Key protease responsible for the removal
CC       of the invariant chain from MHC class II molecules. The bond-
CC       specificity of this proteinase is in part similar to the
CC       specificities of cathepsin L and cathepsin N (By similarity).
CC   -!- CATALYTIC ACTIVITY: Similar to cathepsin L, but with much less
CC       activity on Z-Phe-Arg-|-NHMec, and more activity on the Z-Val-Val-
CC       Arg-|-Xaa compound.
CC   -!- SUBCELLULAR LOCATION: Lysosomal (By similarity).
CC   -!- SIMILARITY: Belongs to the peptidase C1 family.
CC   --------------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   --------------------------------------------------------------------------
DR   EMBL; AY156692; AAO13009.1; -; mRNA.
DR   HSSP; P25774; 1MS6.
DR   SMR; Q8HY81; 115-331.
DR   MEROPS; C01.034; -.
DR   Ensembl; ENSCAFG00000012086; Canis familiaris.
DR   InterPro; IPR000169; Pept_cys_AS.
DR   InterPro; IPR000668; Peptidase_C1.
DR   PANTHER; PTHR12411; Peptidase_C1; 1.
DR   Pfam; PF00112; Peptidase_C1; 1.
DR   PRINTS; PR00705; PAPAIN.
DR   ProDom; PD000158; Peptidase_C1; 1.
DR   SMART; SM00645; Pept_C1; 1.
DR   PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
DR   PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
KW   Glycoprotein; Hydrolase; Lysosome; Protease; Signal; Thiol protease;
KW   Zymogen.
FT   SIGNAL        1     16       Potential.
FT   PROPEP       17    114       Activation peptide (By similarity).
FT   CHAIN       115    331       Cathepsin S.
FT   ACT_SITE    139    139       By similarity.
FT   ACT_SITE    278    278       By similarity.
FT   ACT_SITE    298    298       By similarity.
FT   CARBOHYD    104    104       N-linked (GlcNAc. . .) (Potential).
FT   DISULFID    126    224       By similarity.
FT   DISULFID    136    180       By similarity.
FT   DISULFID    170    213       By similarity.
FT   DISULFID    272    320       By similarity.
```

```
SQ   SEQUENCE   331 AA;  37228 MW;  8E873B7A02105C6A CRC64;

  Query Match             88.5%;  Score 1599;  DB 1;  Length 331;
  Best Local Similarity   87.9%;  Pred. No. 8.2e-125;
  Matches  291;  Conservative   19;  Mismatches   21;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              || || :| :|| ||||:|||||||||:|||||| |||||:||| ||||||||||||
Db          1 MKWLVGLLPLCSYAVAQVHKDPTLDHHWNLWKKTYSKQYKEENEEVARRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              ||||||||||||||||||||||| |||:||| ||||||||||:||:|| |: ||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTGEEVISLMGSLRVPSQWQRNVTYRSNSNQKLPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              |||||||||||||| ||||:|||||||: ||:|||| |||||||||||:| || |||
Db        181 NGGFMTTAFQYIIDNNGIDSEASYPYKAMNGKCRYDSKKRAATCSKYTELPFGSEDALKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              |||||||||| :|| | |||||||||||||||||||||||||:||||:||||||||
Db        241 AVANKGPVSVAIDASHYSFFLYRSGVYYEPSCTQNVNHGVLVVGYGNLNGKDYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              | |||::|||||||| ||||||||:||||||
Db        301 GLNFGDQGYIRMARNSGNHCGIASYPSYPEI 331


RESULT 6
Q6PCU5_RAT
ID   Q6PCU5_RAT  PRELIMINARY;      PRT;   341 AA.
AC   Q6PCU5;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Ctss protein.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Pituitary gland;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Pituitary gland;
RA   Strausberg R.;
RL   Submitted (OCT-2003) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Belongs to the peptidase C1 family.
DR   EMBL; BC059142; AAH59142.1; -; mRNA.
DR   HSSP; P43235; 1AYU.
DR   SMR; Q6PCU5; 124-341.
DR   GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000668; Peptidase_C1.
DR   InterPro; IPR000169; Pept_cys_AS.
DR   Pfam; PF00112; Peptidase_C1; 1.
DR   PRINTS; PR00705; PAPAIN.
DR   ProDom; PD000158; Peptidase_C1; 1.
DR   SMART; SM00645; Pept_C1; 1.
DR   PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
DR   PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
KW   Hydrolase; Protease; Thiol protease.
SQ   SEQUENCE   341 AA;  38573 MW;  833DFC8DD58EC246 CRC64;

  Query Match             77.3%;  Score 1395.5;  DB 2;  Length 341;
  Best Local Similarity   77.5%;  Pred. No. 7.9e-108;
  Matches  255;  Conservative   29;  Mismatches   44;  Indels    1;  Gaps    1;

Qy          4 LVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVMLHN 63
              | | |||| : || :|||||||| |||||: |:||::||| ||||||||||||:||||
Db         13 LFWVPLVCSVTMEQLQRDPTLDHHWDLWKKTHEKEYKDQNEEDVRRLIWEKNLKFIMLHN 72

Qy         64 LEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVDWRE 123
              |||||||||| :||||:|||| |||:  | |||:|  | |: | ||: |: |||||||||
Db         73 LEHSMGMHSYSVGMNHMGDMTPEEVIGYMGSLRIPRHWNRSGTLKSSSNQTLPDSVDWRE 132

Qy        124 KGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCST-EKYGNKGCNG 182
              ||||| |||||||:|||||||||| ||||||||||||||||||||||| |||||||| |
Db        133 KGCVTNVKYQGSCGSCWAFSAVGALEGQLKLKTGKLVSLSAQNLVDCSTEEKYGNKGCGG 192

Qy        183 GFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKEAV 242
              |||| |||||||| ||||:||||||||| || || | ||||||:| |||:| |: |||||
Db        193 GFMTEAFQYIIDNGGIDSEASYPYKAMDEKCHYDPKNRAATCSRYIELPFGDEEALKEAV 252

Qy        243 ANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSWGH 302
              | |||||| :|| | ||||||:|||| :||||:|||||||||||| |:||:|||||||||
Db        253 ATKGPVSVCIDASHSSFFLYQSGVYDDPSCTENVNHGVLVVGYGTLDGKDYWLVKNSWGL 312

Qy        303 NFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              :||::|||||||  |||||||: |||||
Db        313 HFGDQGYIRMARNNKNHCGIASYCSYPEI 341


RESULT 7
Q99M14_MOUSE
ID   Q99M14_MOUSE PRELIMINARY;      PRT;   340 AA.
AC   Q99M14;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Ctss protein.
GN   Name=Ctss;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=CZECH II;
RC   TISSUE=Mammary tumor metastatized to lung. Tumor arose spontaneously;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
```

```
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=CZECH II;
RC   TISSUE=Mammary tumor metastatized to lung. Tumor arose spontaneously;
RA   Strausberg R.;
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Belongs to the peptidase C1 family.
DR   EMBL; BC002125; AAH02125.1; -; mRNA.
DR   HSSP; P25774; 1MS6.
DR   SMR; Q99M14; 123-340.
DR   MGI; MGI:107341; Ctss.
DR   GO; GO:0005615; C:extracellular space; TAS.
DR   GO; GO:0005764; C:lysosome; IDA.
DR   GO; GO:0016020; C:membrane; IDA.
DR   GO; GO:0004218; F:cathepsin S activity; IDA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IDA.
DR   InterPro; IPR000668; Peptidase_C1.
DR   InterPro; IPR000169; Pept_cys_AS.
DR   Pfam; PF00112; Peptidase_C1; 1.
DR   PRINTS; PR00705; PAPAIN.
DR   ProDom; PD000158; Peptidase_C1; 1.
DR   SMART; SM00645; Pept_C1; 1.
DR   PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
DR   PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
KW   Hydrolase; Protease; Thiol protease.
SQ   SEQUENCE   340 AA;  38456 MW;  076502611F319DB7 CRC64;

  Query Match           75.5%;  Score 1363.5;  DB 2;  Length 340;
  Best Local Similarity   75.6%;  Pred. No. 3.7e-105;
  Matches  245;  Conservative   37;  Mismatches   41;  Indels    1;  Gaps    1;

Qy          9 LVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVMLHNLEHSM 68
              |||| |: || :|||||:|| |||||: |:||:|||| |||||||||||:|:|||||:||
Db         17 LVCSVAMEQLQRDPTLDYHWDLWKKTHEKEYKDKNEEEVRRLIWEKNLKFIMIHNLEYSM 76

Qy         69 GMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVDWREKGCVT 128
              |||:| :||| :||||:||::  | :||:| |  : :|::|  || |||:||||||||||
Db         77 GMHTYQVGMNDMGDMTNEEILCRMGALRIPRQSPKTVTFRSYSNRTLPDTVDWREKGCVT 136

Qy        129 EVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCST-EKYGNKGCNGGFMTT 187
              |||||||||||||||||||| ||||||||:|||||||||||  |||||||| ||:||
Db        137 EVKYQGSCGACWAFSAVGALEGQLKLKTGKLISLSAQNLVDCSNEEKYGNKGCGGGYMTE 196

Qy        188 AFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKEAVANKGP 247
              |||||||| ||::|||||||||| || |:|| |||||:| :||:| || |||||| |||
Db        197 AFQYIIDNGGIEADASYPYKAMDEKCHYNSKNRAATCSRYIQLPFGDEDALKEAVATKGP 256

Qy        248 VSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSWGHNFGEE 307
              ||||:|| | ||| |:|||| :|||| |||||||||||| |:||:||||||||| |||::
Db        257 VSVGIDASHSSFFFYKSGVYDDPSCTGNVNHGVLVVGYGTLDGKDYWLVKNSWGLNFGDQ 316

Qy        308 GYIRMARNKGNHCGIASFPSYPEI 331
              |||||||| |||||| |||||
Db        317 GYIRMARNNKNHCGIASDCSYPEI 340


RESULT 8
Q8BSZ5_MOUSE
ID   Q8BSZ5_MOUSE PRELIMINARY;      PRT;   342 AA.
AC   Q8BSZ5;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Mus musculus 18 days pregnant adult female placenta and extra
DE   embryonic tissue cDNA, RIKEN full-length enriched library,
DE   clone:3830425I01 product:cathepsin S, full insert sequence.
GN   Name=Ctss;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Placenta and extra embryonic tissue;
RX   MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA   Carninci P., Hayashizaki Y.;
RT   "High-efficiency full-length cDNA cloning.";
RL   Meth. Enzymol. 303:19-44(1999).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Placenta and extra embryonic tissue;
RX   MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [3]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Placenta and extra embryonic tissue;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
RN   [4]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Placenta and extra embryonic tissue;
RX   MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA   Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA   Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT   "Normalization and subtraction of cap-trapper-selected cDNAs to
RT   prepare full-length cDNA libraries for rapid discovery of new genes.";
RL   Genome Res. 10:1617-1630(2000).
RN   [5]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Placenta and extra embryonic tissue;
RX   MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA   Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA   Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA   Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA   Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA   Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA   Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA   Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT   "RIKEN integrated sequence analysis (RISA) system-384-format
```

```
RT   sequencing pipeline with 384 multicapillary sequencer.";
RL   Genome Res. 10:1757-1771(2000).
RN   [6]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Placenta and extra embryonic tissue;
RA   Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA   Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA   Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA   Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA   Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA   Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA   Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA   Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA   Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA   Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA   Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Belongs to the peptidase C1 family.
DR   EMBL; AK028366; BAC25906.1; -; mRNA.
DR   HSSP; P25774; 1MS6.
DR   SMR; Q8BSZ5; 125-342.
DR   MGI; MGI:107341; Ctss.
DR   GO; GO:0005615; C:extracellular space; TAS.
DR   GO; GO:0005764; C:lysosome; IDA.
DR   GO; GO:0016020; C:membrane; IDA.
DR   GO; GO:0004218; F:cathepsin S activity; IDA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IDA.
DR   InterPro; IPR000668; Peptidase_C1.
DR   InterPro; IPR000169; Pept_cys_AS.
DR   Pfam; PF00112; Peptidase_C1; 1.
DR   PRINTS; PR00705; PAPAIN.
DR   ProDom; PD000158; Peptidase_C1; 1.
DR   SMART; SM00645; Pept_C1; 1.
DR   PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
DR   PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
KW   Hydrolase; Protease; Thiol protease.
SQ   SEQUENCE   342 AA;  38707 MW;  0135655A819BAA85 CRC64;

  Query Match           75.5%;  Score 1363.5;  DB 2;  Length 342;
  Best Local Similarity  75.3%;  Pred. No. 3.7e-105;
  Matches  244;  Conservative   38;  Mismatches   41;  Indels    1;  Gaps    1;

Qy          9 LVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVMLHNLEHSM 68
              |||| |: || :|||||:|| |||||: |:||:|||| ||||||||||||:|:||||:||
Db         19 LVCSVAMEQLQRDPTLDYHWDLWKKTHEKEYKDKNEEEVRRLIWEKNLKFIMIHNLEYSM 78

Qy         69 GMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVDWREKGCVT 128
              |||:| :||| :||||:||:: | :||:| |  : :|::|  || |||:||||||||||
Db         79 GMHTYQVGMNDMGDMTNEEILCRMGALRIPRQSPKTVTFRSYSNRTLPDTVDWREKGCVT 138

Qy        129 EVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCST-EKYGNKGCNGGFMTT 187
              |||||||||||||||||||| |||||||||:|||||||||||  |||||||| ||:||
Db        139 EVKYQGSCGACWAFSAVGALEGQLKLKTGKLISLSAQNLVDCSNEEKYGNKGCGGGYMTE 198

Qy        188 AFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKEAVANKGP 247
              |||||||| ||::|||||||| | || |:|| |||||:| :||:| || |||||| |||
Db        199 AFQYIIDNGGIEADASYPYKATDEKCHYNSKNRAATCSRYIQLPFGDEDALKEAVATKGP 258

Qy        248 VSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSWGHNFGEE 307
              ||||:|| | ||| |:|||| :|||| |||||||||||| |:||:||||||||| |||::
Db        259 VSVGIDASHSSFFFYKSGVYDDPSCTGNVNHGVLVVGYGTLDGKDYWLVKNSWGLNFGDQ 318

Qy        308 GYIRMARNKGNHCGIASFPSYPEI 331
              |||||||  |||||||: |||||
Db        319 GYIRMARNNKNHCGIASYCSYPEI 342


RESULT 9
CATS_MOUSE
ID   CATS_MOUSE     STANDARD;      PRT;   340 AA.
AC   O70370; O54973;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Cathepsin S precursor (EC 3.4.22.27).
GN   Name=Ctss; Synonyms=Cats;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muroidea; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=129/Sv, and BALB/c; TISSUE=Brain;
RA   Doh-ura K.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Spleen;
RA   Rommerskirch W.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   NUCLEOTIDE SEQUENCE OF 144-306, AND TISSUE SPECIFICITY.
RC   STRAIN=C57BL/6; TISSUE=Cartilage;
RX   MEDLINE=99326135; PubMed=10395917;
RA   Soederstroem M., Salminen H., Glumoff V., Kirschke H., Aro H.,
RA   Vuorio E.;
RT   "Cathepsin expression during skeletal development.";
RL   Biochim. Biophys. Acta 1446:35-46(1999).
RN   [4]
RP   NUCLEOTIDE SEQUENCE OF 296-340.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RX   MEDLINE=98184882; PubMed=9516475; DOI=10.1074/jbc.273.13.7691;
RA   Dandoy-Dron F., Guillo F., Benboudjema L., Deslys J.-P., Lasmesas C.,
RA   Dormont D., Tovey M.G., Dron M.;
RT   "Gene expression in scrapie. Cloning of a new scrapie-responsive gene
RT   and the identification of increased levels of seven other mRNA
RT   transcripts.";
RL   J. Biol. Chem. 273:7691-7697(1998).
CC   -!- FUNCTION: Thiol protease. Key protease responsible for the removal
CC       of the invariant chain from MHC class II molecules. The bond-
CC       specificity of this proteinase is in part similar to the
CC       specificities of cathepsin L and cathepsin N.
CC   -!- CATALYTIC ACTIVITY: Similar to cathepsin L, but with much less
CC       activity on Z-Phe-Arg-|-NHMec, and more activity on the Z-Val-Val-
CC       Arg-|-Xaa compound.
CC   -!- SUBCELLULAR LOCATION: Lysosomal.
CC   -!- TISSUE SPECIFICITY: Widely expressed with highest expression found
CC       in non-skeletal tissues. Relatively high levels found in skeletal
CC       tissues.
CC   -!- SIMILARITY: Belongs to the peptidase C1 family.
CC   ----------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ----------------------------------------------------------------------
DR   EMBL; AF051732; AAC05781.1; -; Genomic_DNA.
DR   EMBL; AF051727; AAC05781.1; JOINED; Genomic_DNA.
DR   EMBL; AF051728; AAC05781.1; JOINED; Genomic_DNA.
DR   EMBL; AF051729; AAC05781.1; JOINED; Genomic_DNA.
DR   EMBL; AF051726; AAC05781.1; JOINED; Genomic_DNA.
DR   EMBL; AF051730; AAC05781.1; JOINED; Genomic_DNA.
DR   EMBL; AF051731; AAC05781.1; JOINED; Genomic_DNA.
DR   EMBL; AF038546; AAB94925.1; -; mRNA.
DR   EMBL; AJ002386; CAA05360.1; -; mRNA.
DR   EMBL; Y18466; CAA77184.1; -; mRNA.
DR   EMBL; AJ223208; CAA11182.1; -; mRNA.
DR   PDB; 1M0H; Model; A=1-340.
DR   MEROPS; C01.034; -.
DR   Ensembl; ENSMUSG00000038642; Mus musculus.
DR   MGI; MGI:107341; Ctss.
DR   GO; GO:0005615; C:extracellular space; TAS.
DR   GO; GO:0005764; C:lysosome; IDA.
```

```
DR   GO; GO:0016020; C:membrane; IDA.
DR   GO; GO:0004218; F:cathepsin S activity; IDA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IDA.
DR   InterPro; IPR000169; Pept_cys_AS.
DR   InterPro; IPR000668; Peptidase_C1.
DR   PANTHER; PTHR12411; Peptidase_C1; 1.
DR   Pfam; PF00112; Peptidase_C1; 1.
DR   PRINTS; PR00705; PAPAIN.
DR   ProDom; PD000158; Peptidase_C1; 1.
DR   SMART; SM00645; Pept_C1; 1.
DR   PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
DR   PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
KW   3D-structure; Glycoprotein; Hydrolase; Lysosome; Protease; Signal;
KW   Thiol protease; Zymogen.
FT   SIGNAL        1     17       Potential.
FT   PROPEP       18    122       Activation peptide (Potential).
FT   CHAIN       123    340       Cathepsin S.
FT   ACT_SITE    147    147       By similarity.
FT   ACT_SITE    287    287       By similarity.
FT   ACT_SITE    307    307       By similarity.
FT   CARBOHYD    120    120       N-linked (GlcNAc. . .) (Potential).
FT   DISULFID    134    233       By similarity.
FT   DISULFID    144    189       By similarity.
FT   DISULFID    178    222       By similarity.
FT   DISULFID    281    329       By similarity.
FT   VARIANT     218    218       T -> M.
FT   CONFLICT      1     29       MRAPGHAAIRWLFWMPLVCSVAMEQLQRD -> MAVLDAPG
FT                                VLCGNGATAER (in Ref. 2).
FT   CONFLICT     34     34       Y -> H (in Ref. 2).
FT   CONFLICT     97     97       S -> L (in Ref. 2).
FT   CONFLICT    106    106       S -> P (in Ref. 2).
FT   CONFLICT    146    146       A -> S (in Ref. 3).
SQ   SEQUENCE   340 AA;  38438 MW;  068E61126E2E0C0E CRC64;

  Query Match             74.9%;  Score 1352.5;  DB 1;  Length 340;
  Best Local Similarity   75.0%;  Pred. No. 3e-104;
  Matches  243;  Conservative   37;  Mismatches   43;  Indels    1;  Gaps    1;

Qy          9 LVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVMLHNLEHSM 68
              |||| |: || :|||||:|| |||||: |:||:|||| ||||||||||||:|:||||:||
Db         17 LVCSVAMEQLQRDPTLDYHWDLWKKTHEKEYKDKNEEEVRRLIWEKNLKFIMIHNLEYSM 76

Qy         69 GMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVDWREKGCVT 128
              |||:| :||| :||||:||:   | :||:  |  : :|::|  || |||:||||||||||
Db         77 GMHTYQVGMNDMGDMTNEEISCRMGALRISRQSPKTVTFRSYSNRTLPDTVDWREKGCVT 136

Qy        129 EVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCST-EKYGNKGCNGGFMTT 187
              |||||||||||||||||| |||||||||:|||||||||||  |||||||| ||:||
Db        137 EVKYQGSCGACWAFSAVGALEGQLKLKTGKLISLSAQNLVDCSNEEKYGNKGCGGGYMTE 196

Qy        188 AFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKEAVANKGP 247
              |||||||| ||::|||||||| | || |:|| ||||||:| :||:| || |||||| |||
Db        197 AFQYIIDNGGIEADASYPYKATDEKCHYNSKNRAATCSRYIQLPFGDEDALKEAVATKGP 256

Qy        248 VSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSWGHNFGEE 307
              ||||:|| | ||| |:|||| :|||| |||||||||||| |:||:|||||||| |||::
Db        257 VSVGIDASHSSFFFYKSGVYDDPSCTGNVNHGVLVVGYGTLDGKDYWLVKNSWGLNFGDQ 316

Qy        308 GYIRMARNKGNHCGIASFPSYPEI 331
              |||||||  |||||||: |||||
Db        317 GYIRMARNNKNHCGIASYCSYPEI 340


RESULT 10
CATS_RAT
ID   CATS_RAT       STANDARD;      PRT;   330 AA.
AC   Q02765;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Cathepsin S precursor (EC 3.4.22.27).
GN   Name=Ctss;
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muroidea; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Brain;
RX   MEDLINE=93100327; PubMed=1281481;
RA   Petanceska S., Devi L.;
RT   "Sequence analysis, tissue distribution, and expression of rat
RT   cathepsin S.";
RL   J. Biol. Chem. 267:26038-26043(1992).
CC   -!- FUNCTION: Thiol protease. Key protease responsible for the removal
CC       of the invariant chain from MHC class II molecules. The bond-
CC       specificity of this proteinase is in part similar to the
CC       specificities of cathepsin L and cathepsin N (By similarity). May
CC       be involved in thyroid hormone biosynthesis.
CC   -!- CATALYTIC ACTIVITY: Similar to cathepsin L, but with much less
CC       activity on Z-Phe-Arg-|-NHMec, and more activity on the Z-Val-Val-
CC       Arg-|-Xaa compound.
CC   -!- SUBUNIT: Monomer.
CC   -!- SUBCELLULAR LOCATION: Lysosomal.
CC   -!- TISSUE SPECIFICITY: Highest levels occur in the ileum followed by
CC       spleen, brain, thyroid, ovary and uterus. Low levels are found in
CC       the liver, kidney, jejunum and lung with lowest levels in the
CC       heart.
CC   -!- INDUCTION: By thyroid-stimulating hormone.
CC   -!- SIMILARITY: Belongs to the peptidase C1 family.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; L03201; AAA40994.1; -; mRNA.
DR   PIR; A45087; A45087.
DR   HSSP; P25774; 1MS6.
DR   SMR; Q02765; 113-330.
DR   MEROPS; C01.034; -.
DR   Ensembl; ENSRNOG00000021157; Rattus norvegicus.
DR   RGD; 621513; Ctss.
DR   GO; GO:0004218; F:cathepsin S activity; IDA.
DR   InterPro; IPR000169; Pept_cys_AS.
DR   InterPro; IPR000668; Peptidase_C1.
DR   PANTHER; PTHR12411; Peptidase_C1; 1.
DR   Pfam; PF00112; Peptidase_C1; 1.
DR   PRINTS; PR00705; PAPAIN.
DR   ProDom; PD000158; Peptidase_C1; 1.
DR   SMART; SM00645; Pept_C1; 1.
DR   PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
DR   PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
KW   Glycoprotein; Hydrolase; Lysosome; Protease; Signal; Thiol protease;
KW   Zymogen.
FT   SIGNAL        1     17       Potential.
FT   PROPEP       18    112       Activation peptide.
FT   CHAIN       113    330       Cathepsin S.
FT   ACT_SITE    137    137       By similarity.
FT   ACT_SITE    277    277       By similarity.
FT   ACT_SITE    297    297       By similarity.
FT   CARBOHYD    100    100       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    110    110       N-linked (GlcNAc. . .) (Potential).
FT   DISULFID    124    222       By similarity.
FT   DISULFID    134    179       By similarity.
FT   DISULFID    168    211       By similarity.
FT   DISULFID    271    319       By similarity.
SQ   SEQUENCE   330 AA;  36833 MW;  670E3F08D7749EFE CRC64;

  Query Match             72.0%;  Score 1299.5;  DB 1;  Length 330;
  Best Local Similarity   76.9%;  Pred. No. 7.6e-100;
  Matches  240;  Conservative   28;  Mismatches   41;  Indels    3;  Gaps    3;
```

```
Qy         22 PTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVMLHNLEHSMGMHSYDLGMNHLG 81
              ||||||| |||||  ::  ::||| |||||||||||:|||||||||||||| :||||:|
Db         20 PTLDHHWDLWKKTRMRRNTDQNEEDVRRLIWEKNLKFIMLHNLEHSMGMHSYSVGMNHMG 79

Qy         82 DMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVDWREKGCVTEVKYQGSCGACWA 141
              ||| |||:  | |||:|  | |: | ||: |: ||||||||||||| |||||||:|||
Db         80 DMTPEEVIGYMGSLRIPRPWNRSGTLKSSSNQTLPDSVDWREKGCVTNVKYQGSCGSCWA 139

Qy        142 FSAVGALEAQLKLKTGKLVSLSAQNLVDCST-EKYGNKGCNGGFMTTAFQYIIDNKGIDS 200
              ||| |||| |||||||||||||||||||||| |||||||| ||||| |||||||   |||
Db        140 FSAEGALEGQLKLKTGKLVSLSAQNLVDCSTEEKYGNKGCGGGFMTEAFQYIIDT-SIDS 198

Qy        201 DASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKEAVANKGPVSVGV-DARHPSF 259
              :|||||||| || || | ||||||:| |||:| |: |||||| |||||||: || | ||
Db        199 EASYPYKAMDEKCLYDPKNRAATCSRYIELPFGDEEALKEAVATKGPVSVGIDDASHSSF 258

Qy        260 FLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSWGHNFGEEGYIRMARNKGNH 319
              |||:|||| :||||:|:|||||||||| |:||:||||||||| :||::|||||||  ||
Db        259 FLYQSGVYDDPSCTENMNHGVLVVGYGTLDGKDYWLVKNSWGLHFGDQGYIRMARNNKNH 318

Qy        320 CGIASFPSYPEI 331
              |||||: |||||
Db        319 CGIASYCSYPEI 330


RESULT 11
Q5ZMR6_CHICK
ID   Q5ZMR6_CHICK PRELIMINARY;      PRT;   328 AA.
AC   Q5ZMR6;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Hypothetical protein.
GN   ORFNames=RCJMB04_1f23;
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=CB; TISSUE=Bursa;
RA   Caldwell R.B., Kierzek A.M., Arakawa H., Bezzubov Y., Zaim J.,
RA   Fiedler P., Kutter S., Blagodatski A., Kostovska D., Koter M.,
RA   Plachy J., Carninci P., Hayashizaki Y., Buerstedde J.M.;
RT   "Full-length cDNAs from chicken bursal lymphocytes to facilitate
RT   genefunction analysis.";
RL   Genome Biol. 6:R6-R6(2005).
CC   -!- SIMILARITY: Belongs to the peptidase C1 family.
DR   EMBL; AJ719318; CAG30977.1; -; mRNA.
DR   GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000668; Peptidase_C1.
DR   InterPro; IPR000169; Pept_cys_AS.
DR   Pfam; PF00112; Peptidase_C1; 1.
DR   PRINTS; PR00705; PAPAIN.
DR   ProDom; PD000158; Peptidase_C1; 1.
DR   SMART; SM00645; Pept_C1; 1.
DR   PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
DR   PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
KW   Hydrolase; Hypothetical protein; Protease; Thiol protease.
SQ   SEQUENCE   328 AA;  36397 MW;  C5D007690421FE84 CRC64;

  Query Match             67.4%;  Score 1216.5;  DB 2;  Length 328;
  Best Local Similarity   67.4%;  Pred. No. 6.3e-93;
  Matches  223;  Conservative   36;  Mismatches   69;  Indels    3;  Gaps    2;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              |: | |: :: :      | ||||| || |||| :||:|: : ||  ||  |||||: ||
Db          1 MELLRCMAVLVTLVAVMGHPDPTLDQHWQLWKKAHGKEYRHQAEEGQRRATWEKNLRLVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              |||||||:|:||| |||||:||||||:| :|:: ||||    :  ||:       ||::|
Db         61 LHNLEHSLGLHSYQLGMNHMGDMTSEDVAALLTGLRVPYGHNQTSTYRRRGG--APDAMD 118

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              |||||||||| ||:|||||||||||||||:|||||||||||||||||    ||||||
Db        119 WREKGCVTEVKNQGACGACWAFSAVGALEAQVKLKTGKLVSLSAQNLVDCSM-MYGNKGC 177

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              ||||| |||||||| ||||: |||| | :  |||:   |||||||| ||||  |  ||:
Db        178 GGGFMTRAFQYIIDNNGIDSEESYPYMAQNGTCQYNVSTRAATCSKYVELPYADEAALKD 237

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              |||| ||||| :||  |:|||||||| :| ||| |||||||||| || |::|||||||
Db        238 AVANVGPVSVAIDATQPTFFLYRSGVYDDPRCTQEVNHGVLVVGYGTLNEKDFWLVKNSW 297

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              |  ||: |||||:||  |||||||: |||:|
Db        298 GERFGDGGYIRMSRNHANHCGIASYASYPQI 328


RESULT 12
Q6DJC1_XENTR
ID   Q6DJC1_XENTR PRELIMINARY;      PRT;   333 AA.
AC   Q6DJC1;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Cathepsin S.
GN   Name=ctss-prov;
OS   Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus; Silurana.
OX   NCBI_TaxID=8364;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Whole body;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Whole body;
RA   Klein S., Gerhard D.S.;
RL   Submitted (JUN-2004) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Belongs to the peptidase C1 family.
DR   EMBL; BC075261; AAH75261.1; -; mRNA.
DR   Ensembl; ENSXETG00000008914; Xenopus tropicalis.
DR   GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000169; Pept_cys_AS.
DR   InterPro; IPR000668; Peptidase_C1.
DR   Pfam; PF00112; Peptidase_C1; 1.
DR   PRINTS; PR00705; PAPAIN.
```

```
DR   ProDom; PD000158; Peptidase_C1; 1.
DR   SMART; SM00645; Pept_C1; 1.
DR   PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
DR   PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
KW   Hydrolase; Protease; Thiol protease.
SQ   SEQUENCE   333 AA;  36839 MW;  656E55F8A354407D CRC64;

  Query Match           63.5%;  Score 1147;  DB 2;  Length 333;
  Best Local Similarity   61.2%;  Pred. No. 4e-87;
  Matches  205;  Conservative   56;  Mismatches   68;  Indels    6;  Gaps    3;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              || |: :  |  :|: :   :| ||:|| ||| |: | |::: |:  ||: |||||  |
Db          1 MKSLIFLATVL-AAIVKARINPALDNHWLLWKNTHNKDYEDEIEDLQRRITWEKNLNLVN 59

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRI----LP 116
              :||||:|||||:|:|||||| ||||||: | :: | :| | :|  |: |  |      :|
Db         60 MHNLEYSMGMHTYELGMNHLADMTSEEIKSKLTGLILPPQSERQATFSSQKNSTFGGKVP 119

Qy        117 DSVDWREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYG 176
              ||:|||:||||::|| || ||:|||||||||| || |||||||| |||||||: |||
Db        120 DSIDWRDKGCVSDVKNQGGCGSCWAFSAVGALEGQLMLKTGKLVSLSPQNLVDCSS-KYG 178

Qy        177 NKGCNGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGRED 236
              |||| ||||| ||||:||||||||: ||| ||| || ||   :|:||:||||:  | ||
Db        179 NKGCGGGFMTQAFQYVIDNKGIDSDSYYPYHAMDEKCHYDPTGKASTCAKYTEIVPGTED 238

Qy        237 VLKEAVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLV 296
               ||:|: : ||:|| :|   |||||||||| :|:|:  |||||| ||||:|||:::||:
Db        239 NLKQALGSIGPISVAIDGTRPSFFLYRSGVYSDPTCSHEVNHGVLAVGYGNLNGQDFWLL 298

Qy        297 KNSWGHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              |||||  :|::||:|:|||||| ||:||:  ||||
Db        299 KNSWGTKYGDQGYVRIARNKGNLCGVASYTCYPEI 333


RESULT 13
Q6DE57_XENLA
ID   Q6DE57_XENLA PRELIMINARY;      PRT;   320 AA.
AC   Q6DE57;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Ctss-prov protein.
GN   Name=ctss-prov;
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Spleen;
RX   MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
RA   Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA   Richardson P.;
RT   "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT   initiative.";
RL   Dev. Dyn. 225:384-391(2002).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Spleen;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Spleen;
RA   Klein S., Gerhard D.S.;
RL   Submitted (JUL-2004) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Belongs to the peptidase C1 family.
DR   EMBL; BC077285; AAH77285.1; -; mRNA.
DR   GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000668; Peptidase_C1.
DR   InterPro; IPR000169; Pept_cys_AS.
DR   Pfam; PF00112; Peptidase_C1; 1.
DR   PRINTS; PR00705; PAPAIN.
DR   ProDom; PD000158; Peptidase_C1; 1.
DR   SMART; SM00645; Pept_C1; 1.
DR   PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
DR   PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
KW   Hydrolase; Protease; Thiol protease.
SQ   SEQUENCE   320 AA;  35654 MW;  1D7C3DC9ADC433FA CRC64;

  Query Match           62.3%;  Score 1124.5;  DB 2;  Length 320;
  Best Local Similarity   63.2%;  Pred. No. 2.9e-85;
  Matches  199;  Conservative   50;  Mismatches   61;  Indels    5;  Gaps    2;

Qy         21 DPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVMLHNLEHSMGMHSYDLGMNHL 80
              || ||:|| |||  : |:|::::|: :||: ||||| | :||||:|||||:|:||||||
Db          7 DPALDNHWLLWKNKHTKEYEDESEDLLRRITWEKNLNTVNMHNLEYSMGMHTYELGMNHL 66

Qy         81 GDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRIL----PDSVDWREKGCVTEVKYQGSC 136
              ||||||: | |: | :|   :|  |: |  |  |    |||:||||||||:||| || |
Db         67 ADMTSEEIKSKMTGLILPPHSERKATFSSQKNSTLGGKVPDSIDWREKGCVSEVKNQGGC 126

Qy        137 GACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGCNGGFMTTAFQYIIDNK 196
              |:||||||||||| || ||||:|||| |||||||: |||||||:||||| ||||:|||
Db        127 GSCWAFSAVGALEGQLMLKTGKIVSLSPQNLVDCSS-KYGNKGCSGGFMTRAFQYVIDNN 185

Qy        197 GIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKEAVANKGPVSVGVDARH 256
              |||||  ||| ||| || |:   :|::| || |:  | || ||:|: | ||:|| :|
Db        186 GIDSDTYYPYHAMDEKCHYELAGKASSCVKYREIVPGTEDNLKQALGNIGPISVAIDGTR 245

Qy        257 PSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSWGHNFGEEGYIRMARNK 316
              |:||||:|||| :|||:| |||||| |||| |||:::||:|||||  :|::||:|:||||
Db        246 PTFFLYKSGVYSDPSCSQEVNHGVLAVGYGTLNGQDFWLLKNSWGTKYGDQGYVRIARNK 305

Qy        317 GNHCGIASFPSYPEI 331
              | ||:||: |||||
Db        306 ENLCGVASYTSYPEI 320


RESULT 14
Q7T0S4_XENLA
ID   Q7T0S4_XENLA PRELIMINARY;      PRT;   333 AA.
AC   Q7T0S4;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Ctss-prov protein.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus; Xenopus.
```

```
OX   NCBI_TaxID=8355;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Spleen;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Spleen;
RX   MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
RA   Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA   Richardson P.;
RT   "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT   initiative.";
RL   Dev. Dyn. 225:384-391(2002).
RN   [3]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Spleen;
RA   Klein S., Strausberg R.;
RL   Submitted (AUG-2003) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Belongs to the peptidase C1 family.
DR   EMBL; BC056059; AAH56059.1; -; mRNA.
DR   HSSP; P25774; 1NQC.
DR   MEROPS; C01.034; -.
DR   GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000668; Peptidase_C1.
DR   InterPro; IPR000169; Pept_cys_AS.
DR   Pfam; PF00112; Peptidase_C1; 1.
DR   PRINTS; PR00705; PAPAIN.
DR   ProDom; PD000158; Peptidase_C1; 1.
DR   SMART; SM00645; Pept_C1; 1.
DR   PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
DR   PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
KW   Hydrolase; Protease; Thiol protease.
SQ   SEQUENCE   333 AA;  36943 MW;  655534D5084059FA CRC64;

  Query Match             61.4%;  Score 1109.5;  DB 2;  Length 333;
  Best Local Similarity   58.9%;  Pred. No. 5.4e-84;
  Matches  196;  Conservative   61;  Mismatches   71;  Indels    5;  Gaps    2;

Qy          3 RLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVMLH 62
              |   :| :  :|  :   :| ||:|| ||| |: |:|::: |:  ||: ||||| || :|
Db          2 RSFILLAIALTATVKARINPALDNHWLLWKNTHSKEYEDETEDLQRRITWEKNLDFVNMH 61

Qy         63 NLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRI----LPDS 118
              |||:|||||:|:|||||| ||||||: | :: | :|   :|   : | |      : ||
Db         62 NLEYSMGMHTYELGMNHLADMTSEEMKSKLTGLILPPHSERKAKFSSQRNGTFGGKVRDS 121

Qy        119 VDWREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNK 178
              :|||:||||::|| || ||:|||||||||| || |||||||||| ||||||:: |||||
Db        122 IDWRDKGCVSDVKNQGGCGSCWAFSAVGALEGQLMLKTGKLVSLSPQNLVDCAS-KYGNK 180

Qy        179 GCNGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVL 238
              ||:|||||:||||:||| ||||| ||| ||| || |:   :|::| ||||:  | || |
Db        181 GCSGGFMTSAFQYVIDNNGIDSDSYYPYHAMDEKCHYELAGKASSCVKYTEIVPGTEDNL 240

Qy        239 KEAVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKN 298
              |:|:   ||:|| :|   |:||||:|||| :|||:| |||||| :||| |||:::||:||
Db        241 KQALGTIGPISVAIDGTRPTFFLYKSGVYSDPSCSQEVNHGVLAIGYGTLNGQDFWLLKN 300

Qy        299 SWGHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              |||  :|::|::|:|||||| ||:||: |||||
Db        301 SWGTYYGDKGFVRIARNKGNLCGVASYTSYPEI 333


RESULT 15
Q7T183_9CICH
ID   Q7T183_9CICH PRELIMINARY;      PRT;   334 AA.
AC   Q7T183;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Cathepsin.
OS   Paralabidochromis chilotes.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Labroidei;
OC   Cichlidae; African cichlids; Pseudocrenilabrinae; Haplochromini;
OC   Paralabidochromis.
OX   NCBI_TaxID=77306;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Head;
RX   MEDLINE=22871689; PubMed=14507309;
RX   DOI=10.1046/j.1365-3083.2003.01322.x;
RA   Uinuk-Ool T.S., Takezaki N., Kuroda N., Figueroa F., Sato A.,
RA   Samonte I.E., Mayer W.E., Klein J.;
RT   "Phylogeny of antigen-processing enzymes: cathepsins of a
RT   cephalochordate, an agnathan and a bony fish.";
RL   Scand. J. Immunol. 58:436-448(2003).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Head;
RA   Uinuk-ool T.S., Takezaki N., Kuroda N., Figueroa F., Sato A.,
RA   Samonte I.E., Mayer W.E., Klein J.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Belongs to the peptidase C1 family.
DR   EMBL; AY333300; AAQ01147.1; -; mRNA.
DR   HSSP; P25774; 1NQC.
DR   MEROPS; C01.034; -.
DR   GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000668; Peptidase_C1.
DR   InterPro; IPR000169; Pept_cys_AS.
DR   Pfam; PF00112; Peptidase_C1; 1.
DR   PRINTS; PR00705; PAPAIN.
DR   ProDom; PD000158; Peptidase_C1; 1.
DR   SMART; SM00645; Pept_C1; 1.
DR   PROSITE; PS00640; THIOL_PROTEASE_ASN; 1.
DR   PROSITE; PS00139; THIOL_PROTEASE_CYS; 1.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
KW   Hydrolase; Protease; Thiol protease.
SQ   SEQUENCE   334 AA;  36543 MW;  A227955CB423E537 CRC64;

  Query Match             60.1%;  Score 1086;  DB 2;  Length 334;
  Best Local Similarity   60.9%;  Pred. No. 4.9e-82;
  Matches  199;  Conservative   40;  Mismatches   84;  Indels    4;  Gaps    3;

Qy          4 LVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVMLHN 63
              |: :: :|:    |    : ||| || ||||:|| ||   | | || :|  ||| : :||
Db          9 LLLLISLCAGVAAMF--ESTLDAHWELWKKTHGKSYKNDVENAHRRELWGNNLKMITVHN 66

Qy         64 LEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNIT-YKSNPNRILPDSVDWR 122
              || |||:|:|:|||||:||:| ||:|   :||  |:  ||  : :       :||::|||
Db         67 LEASMGLHTYELGMNHMGDLTEEEIMQFFASLTPPTDIQRAPSPFAGASGSGIPDTMDWR 126
```

```
Qy        123 EKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGCNG 182
              ||||||:|| ||:||:|||||| |||| ||   ||||| || ||||||| |||| ||||
Db        127 EKGCVTKVKMQGACGSCWAFSAAGALEGQLAKSTGKLVDLSPQNLVDCS-GKYGNHGCNG 185

Qy        183 GFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKEAV 242
              |||| ||||:||| |||||||||||    | :| |:   ||| || |  || | |: ||: :
Db        186 GFMTRAFQYVIDNHGIDSDASYPYIGRDDQCHYNPATRAANCSSYQFLPEGDENALKQGL 245

Qy        243 ANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSWGH 302
              |  ||:|| :||| | |  |||||| :||||| |||||| |||| |||::|||||||||
Db        246 ATVGPISVAIDARRPRFSFYRSGVYNDPSCTQKVNHGVLAVGYGTLNGQDYWLVKNSWGT 305

Qy        303 NFGEEGYIRMARNKGNHCGIASFPSYP 329
               ||::|||||||| || |||| :| ||
Db        306 TFGDQGYIRMARNTGNQCGIALYPCYP 332
```

Search completed: January 10, 2006, 09:35:42
Job time : 165 secs

OM protein - protein search, using sw model

Run on: January 10, 2006, 09:36:39 ; Search time 50 Seconds
(without alignments)
56.228 Million cell updates/sec

Title: US-10-646-470-1
Perfect score: 1806
Sequence: 1 MKRLVCVLLVCSSAVAQLHK.....MARNKGNHCGIASFPSYPEI 331

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 61141 seqs, 8493638 residues

Total number of hits satisfying chosen parameters: 61141

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA_New:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1800	99.7	331	6	US-10-995-561-977	Sequence 977, App
2	1800	99.7	331	6	US-10-995-561-978	Sequence 978, App
3	1129	62.5	220	6	US-10-995-561-976	Sequence 976, App
4	849	47.0	334	6	US-10-131-826A-12	Sequence 12, Appl
5	840.5	46.5	333	6	US-10-821-234-1672	Sequence 1672, Ap
6	536.5	29.7	335	6	US-10-995-561-866	Sequence 866, App
7	408	22.6	515	6	US-10-821-234-1211	Sequence 1211, Ap
8	346.5	19.2	262	7	US-11-183-914-2	Sequence 2, Appli
9	324	17.9	488	6	US-10-821-234-1000	Sequence 1000, Ap
10	318.5	17.6	320	7	US-11-102-883-20	Sequence 20, Appl
11	309	17.1	339	6	US-10-878-556A-18	Sequence 18, Appl
12	308	17.1	339	6	US-10-995-561-681	Sequence 681, App
13	308	17.1	339	6	US-10-995-561-682	Sequence 682, App
14	308	17.1	339	6	US-10-995-561-684	Sequence 684, App
15	308	17.1	339	6	US-10-995-561-685	Sequence 685, App
16	308	17.1	339	6	US-10-995-561-686	Sequence 686, App
17	308	17.1	339	6	US-10-995-561-687	Sequence 687, App
18	308	17.1	339	7	US-11-186-284-43	Sequence 43, Appl
19	308	17.1	344	6	US-10-821-234-923	Sequence 923, App
20	264.5	14.6	362	7	US-11-102-883-8	Sequence 8, Appli
21	261.5	14.5	351	7	US-11-102-883-28	Sequence 28, Appl
22	260.5	14.4	467	7	US-11-037-243-63	Sequence 63, Appl
23	99.5	5.5	1767	7	US-11-052-554A-372	Sequence 372, App
24	85.5	4.7	437	6	US-10-873-427A-33	Sequence 33, Appl
25	83.5	4.6	693	7	US-11-167-856-2	Sequence 2, Appli
26	79.5	4.4	557	7	US-11-191-374-4	Sequence 4, Appli
27	79.5	4.4	557	7	US-11-191-375-4	Sequence 4, Appli
28	79.5	4.4	557	7	US-11-191-588-4	Sequence 4, Appli
29	78.5	4.3	472	7	US-11-114-301-2	Sequence 2, Appli
30	77.5	4.3	334	7	US-11-016-564-1	Sequence 1, Appli
31	76	4.2	330	6	US-10-485-517-415	Sequence 415, App
32	76	4.2	338	6	US-10-467-657-136	Sequence 136, App
33	76	4.2	338	6	US-10-467-657-6798	Sequence 6798, Ap
34	76	4.2	655	7	US-11-045-802-29	Sequence 29, Appl
35	76	4.2	690	7	US-11-052-554A-232	Sequence 232, App
36	74.5	4.1	621	7	US-11-184-380-26	Sequence 26, Appl
37	74	4.1	1474	6	US-10-995-561-873	Sequence 873, App
38	73.5	4.1	232	6	US-10-821-234-970	Sequence 970, App
39	73.5	4.1	2367	7	US-11-051-453-42	Sequence 42, Appl
40	72.5	4.0	590	6	US-10-632-150-52	Sequence 52, Appl
41	72.5	4.0	590	7	US-11-073-457-52	Sequence 52, Appl
42	72.5	4.0	590	7	US-11-073-460-52	Sequence 52, Appl
43	72.5	4.0	755	7	US-11-037-243-82	Sequence 82, Appl
44	72.5	4.0	985	7	US-11-113-424-61	Sequence 61, Appl
45	71.5	4.0	727	6	US-10-995-561-864	Sequence 864, App

ALIGNMENTS

RESULT 1
US-10-995-561-977
; Sequence 977, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 977
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-977

Query Match 99.7%; Score 1800; DB 6; Length 331;
Best Local Similarity 99.7%; Pred. No. 3.2e-160;
Matches 330; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              |||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
```

```
RESULT 2
US-10-995-561-978
; Sequence 978, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 978
;   LENGTH: 331
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-995-561-978

  Query Match             99.7%;  Score 1800;  DB 6;  Length 331;
  Best Local Similarity   99.7%;  Pred. No. 3.2e-160;
  Matches  330;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              |||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331


RESULT 3
US-10-995-561-976
; Sequence 976, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 976
;   LENGTH: 220
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-995-561-976

  Query Match             62.5%;  Score 1129;  DB 6;  Length 220;
  Best Local Similarity   100.0%;  Pred. No. 4.8e-98;
  Matches  209;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAM 209
              |||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAM 209


RESULT 4
US-10-131-826A-12
; Sequence 12, Application US/10131826A
; Publication No. US20050245730A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C128
;  CURRENT APPLICATION NUMBER: US/10/131,826A
;  CURRENT FILING DATE:  2002-04-24
;  PRIOR APPLICATION NUMBER: 60/049911
;  PRIOR FILING DATE: 1997-06-18
;  PRIOR APPLICATION NUMBER: 60/056974
;  PRIOR FILING DATE: 1997-08-26
;  PRIOR APPLICATION NUMBER: 60/059113
;  PRIOR FILING DATE: 1997-09-17
;  PRIOR APPLICATION NUMBER: 60/059115
;  PRIOR FILING DATE: 1997-09-17
;  PRIOR APPLICATION NUMBER: 60/059117
;  PRIOR FILING DATE: 1997-09-17
;  PRIOR APPLICATION NUMBER: 60/059122
;  PRIOR FILING DATE: 1997-09-17
;  PRIOR APPLICATION NUMBER: 60/059184
;  PRIOR FILING DATE: 1997-09-17
;  PRIOR APPLICATION NUMBER: 60/059263
;  PRIOR FILING DATE: 1997-09-18
;  PRIOR APPLICATION NUMBER: 60/059352
;  PRIOR FILING DATE: 1997-09-19
;  PRIOR APPLICATION NUMBER: 60/059588
;  PRIOR FILING DATE: 1997-09-19
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 12
;   LENGTH: 334
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-131-826A-12
```

```
  Query Match           47.0%;  Score 849;  DB 6;  Length 334;
  Best Local Similarity  49.8%;  Pred. No. 9.2e-72;
  Matches  166;  Conservative   55;  Mismatches  102;  Indels   10;  Gaps    6;

Qy          4 LVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVMLHN 63
              |     | :||| :   |  ||  |: || |: : |   |||  || :||||:| : |||
Db          7 LAAFCLGIASAVPKF--DQNLDTKWYQWKATHRRLY-GANEEGWRRAVWEKNMKMIELHN 63

Qy         64 LEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVDWRE 123
              |:| | | : : ||  ||||:||   :|   |   ::::   ::       || |||||:
Db         64 GEYSQGKHGFTMAMNAFGDMTNEEFRQMMGCFR-NQKFRKGKVFREPLFLDLPKSVDWRK 122

Qy        124 KGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGCNGG 183
              || || || |  ||:|||||| |||| |:  ||||||||| |||||||  : ||:|||||
Db        123 KGYVTPVKNQKQCGSCWAFSATGALEGQMFRKTGKLVSLSEQNLVDCSRPQ-GNQGCNGG 181

Qy        184 FMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKEAVA 243
              ||  |||| :| |:||: |||| |:|  |:|  :   |  : :| :  |:|  | :|||
Db        182 FMARAFQYVKENGGLDSEESYPYVAVDEICKYRPENSVANDTGFTVVAPGKEKALMKAVA 241

Qy        244 NKGPVSVGVDARHPSFFLYRSGVYYEPSC-TQNVNHGVLVVGYG----DLNGKEYWLVKN 298
              ||:|| :|| | ||  |:||:|:|| | ::|::|||||||||    : |  :||||||
Db        242 TVGPISVAMDAGHSSFQFYKSGIYFEPDCSSKNLDHGVLVVGYGFEGANSNNSKYWLVKN 301

Qy        299 SWGHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              |||  :|  ||:::|::| ||||||:  ||| :
Db        302 SWGPEWGSNGYVKIAKDKNNHCGIATAASYPNV 334


RESULT 5
US-10-821-234-1672
; Sequence 1672, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
;  APPLICANT: Labat, Ivan
;  APPLICANT:  Stache-Crain, Birgit
;  APPLICANT:  Andarmani, Susan
;  APPLICANT:  Tang, Y. Tom
;  TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
;  FILE REFERENCE: 821A
;  CURRENT APPLICATION NUMBER: US/10/821,234
;  CURRENT FILING DATE:  2004-04-07
;  PRIOR APPLICATION NUMBER: US 60/462,047
;  PRIOR FILING DATE: 2003-04-07
;  NUMBER OF SEQ ID NOS: 1704
;  SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 1672
;   LENGTH: 333
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-821-234-1672

  Query Match           46.5%;  Score 840.5;  DB 6;  Length 333;
  Best Local Similarity  48.5%;  Pred. No. 5.7e-71;
  Matches  164;  Conservative   57;  Mismatches   98;  Indels   19;  Gaps    8;

Qy          4 LVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVMLHN 63
              |:        | | |  | :|:  |  ||  : : |   |||  || :||||:| : |||
Db          5 LILAAFCLGIASATLTFDHSLEAQWTKWKAMHNRLY-GMNEEGWRRAVWEKNMKMIELHN 63

Qy         64 LEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSL--RVPSQ---WQRNITYKSNPNRILPDS 118
              |:  | ||: : ||  |||||||   :|:    | | :   :|  : |::      | |
Db         64 QEYREGKHSFTMAMNAFGDMTSEEFRQVMNGFQNRKPRKGKVFQEPLFYEA------PRS 117

Qy        119 VDWREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNK 178
              ||||||| || || || ||:|||||| |||| |:  |||:|:||| |||||||  : ||:
Db        118 VDWREKGYVTPVKNQGQCGSCWAFSATGALEGQMFRKTGRLISLSEQNLVDCSGPQ-GNE 176

Qy        179 GCNGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVL 238
              ||||| |  ||||: || |:||: ||||:| :  |:|: ||   |  : : ::|  :|  |
Db        177 GCNGGLMDYAFQYVQDNGGLDSEESYPYEATEESCKYNPKYSVANDTGFVDIP-KQEKAL 235

Qy        239 KEAVANKGPVSVGVDARHPSFFLYRSGVYYEPSC-TQNVNHGVLVVGYG----DLNGKEY 293
              :|||  ||:|| :|| | ||  |: |:|:||| | :::::|||||||||    : :  :|
Db        236 MKAVATVGPISVAIDAGHESFLFYKEGIYFEPDCSSEDMDHGVLVVGYGFESTESDNNKY 295

Qy        294 WLVKNSWGHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              ||||||||  :|  ||::||::: |||||||  ||| :
Db        296 WLVKNSWGEEWGMGGYVKMAKDRRNHCGIASAASYPTV 333


RESULT 6
US-10-995-561-866
; Sequence 866, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 866
;   LENGTH: 335
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-995-561-866

  Query Match           29.7%;  Score 536.5;  DB 6;  Length 335;
  Best Local Similarity  37.9%;  Pred. No. 1.1e-42;
  Matches  125;  Conservative   49;  Mismatches  129;  Indels   27;  Gaps   12;

Qy         10 VCSSA---VAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRL-IWEKNLKFVMLHNLE 65
              || :|   |  | |      |:  |   :  | |    ||   ||  :  | : :  ||
Db         19 VCGAAELCVNSLEK-----FHFKSWMSKHRKTY--STEEYHHRLQTFASNWRKINAHN-- 69

Qy         66 HSMGMHSYDLGMNHLGDMTSEEVMS--LMSSLRVPSQWQRNITYKSNPNRILPDSVDWRE 123
                 | |:: : :|   ||:  |:     | |  :  |  : |     : |     | |||||:
Db         70 --NGNHTFKMALNQFSDMSFAEIKHKYLWSEPQNCSATKSNYLRGTGP---YPPSVDWRK 124

Qy        124 KG-CVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGCNG 182
              ||  |: || ||:||:|| ||  ||||: : : |||::||: | ||||: : : | || |
Db        125 KGNFVSPVKNQGACGSCWTFSTTGALESAIAIATGKMLSLAEQQLVDCA-QDFNNHGCQG 183

Qy        183 GFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKEAV 242
              |  : ||:||: ||||  : :|||:  |  |::              :   |: : |||
Db        184 GLPSQAFEYILYNKGIMGEDTYPYQGKDGYCKFQPGKAIGFVKDVANITIYDEEAMVEAV 243

Qy        243 ANKGPVSVGVDARHPSFFLYRSGVYYEPSC---TQNVNHGVLVVGYGDLNGKEYWLVKNS 299
              |   |||   :      | :||:|:|    ||       ||| || ||||: ||  ||:||||
Db        244 ALYNPVSFAFEVTQ-DFMMYRTGIYSSTSCHKTPDKVNHAVLAVGYGEKNGIPYWIVKNS 302

Qy        300 WGHNFGEEGYIRMARNKGNHCGIASFPSYP 329
              ||  :|  ||  : | | | ||:|:  |||
Db        303 WGPQWGMNGYFLIERGK-NMCGLAACASYP 331


RESULT 7
US-10-821-234-1211
; Sequence 1211, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
;  APPLICANT: Labat, Ivan
;  APPLICANT:  Stache-Crain, Birgit
;  APPLICANT:  Andarmani, Susan
;  APPLICANT:  Tang, Y. Tom
;  TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
;  FILE REFERENCE: 821A
;  CURRENT APPLICATION NUMBER: US/10/821,234
;  CURRENT FILING DATE:  2004-04-07
;  PRIOR APPLICATION NUMBER: US 60/462,047
;  PRIOR FILING DATE: 2003-04-07
;  NUMBER OF SEQ ID NOS: 1704
```

```
;  SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 1211
;   LENGTH: 515
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-821-234-1211

  Query Match             22.6%;  Score 408;  DB 6;  Length 515;
  Best Local Similarity   31.1%;  Pred. No. 1.7e-30;
  Matches  103;  Conservative   62;  Mismatches  136;  Indels   30;  Gaps   10;

Qy         12 SSAVAQLHKDP-------TLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVMLHNL 64
              || ::  |::||        :   :  :  || : |: | |||  ||        | ::|: 
Db        195 SSVISLLNEDPLSQDLPVKMASIFKNFVITYNRTYESK-EEARWRL-------SVFVNNM 246

Qy         65 EHSMGMHSYDLGMNHLG-----DMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSV 119
                :  : : | |    |     |:| ||  ::  :  :  :    :    :   : |
Db        247 VRAQKIQALDRGTAQYGVTKFSDLTEEEFRTIYLNTLLRKEPGNKMKQAKSVGDLAPPEW 306

Qy        120 DWREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKG 179
              ||| || ||:|| || ||:|||||  | :| |  |  | |:||| | |:||      :|
Db        307 DWRSKGAVTKVKDQGMCGSCWAFSVTGNVEGQWFLNQGTLLSLSEQELLDCDKM---DKA 363

Qy        180 CNGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLK 239
              | ||  : |:  | :  |:::: | |:     | : ::      :   ||    | |
Db        364 CMGGLPSNAYSAIKNLGGLETEDDYSYQGHMQSCNFSAEKAKVYINDSVELSQ-NEQKLA 422

Qy        240 EAVANKGPVSVGVDARHPSFFLYRSGVY--YEPSCTQ-NVNHGVLVVGYGDLNGKEYWLV 296
              :| :|||:|| ::|    |  || |:     | |:   ::| ||:|||: :   :| :
Db        423 AWLAKRGPISVAINAFGMQF--YRHGISRPLRPLCSPWLIDHAVLLVGYGNRSDVPFWAI 480

Qy        297 KNSWGHNFGEEGYIRMARNKGNHCGIASFPS 327
              ||||| ::||:||  : |  |  ||: :  |
Db        481 KNSWGTDWGEKGYYYLHRGSG-ACGVNTMAS 510


RESULT 8
US-11-183-914-2
; Sequence 2, Application US/11183914
; Publication No. US20050282214A1
;  GENERAL INFORMATION:
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Yue, Henry
;    APPLICANT:  Guegler, Karl J.
;    APPLICANT:  Corley, Neil C.
;    APPLICANT:  Tang, Tom Y.
;    APPLICANT:  Shah, Purvi
;    TITLE OF INVENTION:  HUMAN PROTEASE MOLECULES
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Dr.
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/11/183,914
;      FILING DATE:  19-JULY-2005
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/008,271
;      FILING DATE:  16-Jan-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  <Unknown>
;      FILING DATE:  <Unknown>
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mohan-Peterson, Sheela
;      REGISTRATION NUMBER:  41,201
;      REFERENCE/DOCKET NUMBER:  PF-0458 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-855-0555
;      TELEFAX:  650-845-4166
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  262 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  TLYMNOT02
;      CLONE:  447484
US-11-183-914-2

  Query Match             19.2%;  Score 346.5;  DB 7;  Length 262;
  Best Local Similarity   32.0%;  Pred. No. 3.7e-25;
  Matches   80;  Conservative   44;  Mismatches   77;  Indels   49;  Gaps   10;

Qy        111 PNRILPDSVDWRE-KGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVD 169
              |   :| | |||:  | :: :| | :|  ||| :| | :|   ::      | :| | |:|
Db         10 PEESVPFSCDWRKVAGAISPIKDQKNCNCCWAMAAAGNIETLWRISFWDFVDVSVQELLD 69

Qy        170 CSTEKYGNKGCNGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATC--SKY 227
              |   : |: ||:|||:  ||  :::| |: |:  ||::          | ||  |   ||
Db         70 CG--RCGD-GCHGGFVWDAFITVLNNSGLASEKDYPFQ---------GKVRAHRCHPKKY 117

Qy        228 TELPY--------GREDVLKEAVANKGPVSVGVDARHPSFFLYRSGVYYEPSCT---QNV 276
               :: :          |  : : :|  ||::|  :: :     ||| ||       |   | |
Db        118 QKVAWIQDFIMLQNNEHRIAQYLATYGPITVTINMK--PLQLYRKGVIKATPTTCDPQLV 175

Qy        277 NHGVLVVGYGDLNGKE-------------------YWLVKNSWGHNFGEEGYIRMARNK 316
              :| ||:||:| :  :|                   ||::|||||  :||:|| |: |
Db        176 DHSVLLVGFGSVKSEEGIWAETVSSQSQPQPPHPTPYWILKNSWGAQWGEKGYFRLHRG- 234

Qy        317 GNHCGIASFP 326
              | |||  ||
Db        235 SNTCGITKFP 244


RESULT 9
US-10-821-234-1000
; Sequence 1000, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
;  APPLICANT: Labat, Ivan
;  APPLICANT:  Stache-Crain, Birgit
;  APPLICANT:  Andarmani, Susan
;  APPLICANT:  Tang, Y. Tom
;  TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
;  FILE REFERENCE: 821A
;  CURRENT APPLICATION NUMBER: US/10/821,234
;  CURRENT FILING DATE:  2004-04-07
;  PRIOR APPLICATION NUMBER: US 60/462,047
;  PRIOR FILING DATE: 2003-04-07
;  NUMBER OF SEQ ID NOS: 1704
;  SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 1000
;   LENGTH: 488
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-821-234-1000

  Query Match             17.9%;  Score 324;  DB 6;  Length 488;
  Best Local Similarity   32.0%;  Pred. No. 1e-22;
  Matches   90;  Conservative   47;  Mismatches   82;  Indels   62;  Gaps   16;

Qy         80 LGDMTSEEVMSLMSSLRVP--------SQWQRNITYKSNPNRILPDSVDWREK---GCVT 128
              ||||      |   | ::|        :: |: | :        || | |||       |:
Db        223 LGDMIRR---SGGHSRKIPRPKPAPLTAEIQQKILH-------LPTSWDWRNVHGINFVS 272

Qy        129 EVKYQGSCGACWAFSAVGALEAQLKLKTGKLVS--LSAQNLVDCSTEKYGNKGCNGGFMT 186
```

```
                |: | |||:|::|:::| |||:::: |    :  || | :| ||  :|  :|| ||
Db        273 PVRNQASCGSCYSFASMGMLEARIRILTNNSQTPILSPQEVVSCS--QYA-QGCEGG--- 326

Qy        187 TAFQYIIDNK-----GIDSDASYPYKAMDLKCQ--------YDSKYRAATCSKYTELPYG 233
              | |:|  |     |:  :| :||   |  |:        | |:|       |    ||
Db        327 --FPYLIAGKYAQDFGLVEEACFPYTGTDSPCKMKEDCFRYYSSEYH------YVGGFYG 378

Qy        234 --REDVLKEAVANKGPVSVGVDARHPSFFLYRSGVYYEPSCT------QNVNHGVLVVGY 285
                | ::|  : : ||::|  :  :  |  |: |:|:           :  || ||:|||
Db        379 GCNEALMKLELVHHGPMAVAFEV-YDDFLHYKKGIYHHTGLRDPFNPFELTNHAVLLVGY 437

Qy        286 G--DLNGKEYWLVKNSWGHNFGEEGYIRMARNKGNHCGIAS 324
              |    :| :||:|||||  :|| || |: |   : | | |
Db        438 GTDSASGMDYWIVKNSWGTGWGENGYFRIRRGT-DECAIES 477


RESULT 10
US-11-102-883-20
; Sequence 20, Application US/11102883
; Publication No. US20050281816A1
; GENERAL INFORMATION:
;  APPLICANT: Lamping, Norbert
;  APPLICANT:  Crameri, Reto
;  APPLICANT:  Fluckiger, Sabina
;  APPLICANT:  Daigle, Isabelle
;  TITLE OF INVENTION: Modular Antigen Transporter Molecules (MAT Molecules) for
;  TITLE OF INVENTION:  Modulating Immune Reactions,  Associated Constructs, Methods and
;  TITLE OF INVENTION:  Uses Thereof
;  FILE REFERENCE: 03100234pa
;  CURRENT APPLICATION NUMBER: US/11/102,883
;  CURRENT FILING DATE:  2005-04-11
;  PRIOR APPLICATION NUMBER: EP02022774.0
;  PRIOR FILING DATE: 2002-10-11
;  PRIOR APPLICATION NUMBER: PCT/EP2003/011190
;  PRIOR FILING DATE: 2003-10-09
;  NUMBER OF SEQ ID NOS: 44
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 20
;   LENGTH: 320
;   TYPE: PRT
;   ORGANISM: Dermatophagoides pteronyssinus
US-11-102-883-20

  Query Match             17.6%;  Score 318.5;  DB 7;  Length 320;
  Best Local Similarity   28.0%;  Pred. No. 1.9e-22;
  Matches   91;  Conservative   66;  Mismatches  129;  Indels   39;  Gaps   12;

Qy          7 VLLVCS-SAVAQLHKDPTLDHHWHLWKKTYGKQYKE-KNEEAVRRLIWEKNLKFVMLHNL 64
              || : |  |:: ::  |:    :  :|| : | |   ::||| |:   | ::|:|
Db          4 VLAIASLLALSAVYARPSSIKTFEEYKKAFNKSYATFEDEEAARKNFLE-SVKYV----- 57

Qy         65 EHSMGMHSYDLGMNHLGDMTSEE-----VMSLMSSLRVPSQWQRNI-TYKSNPNRILPDS 118
                    |    :||| |:: :|     :||  :   : :|:  |  |   : |   |
Db         58 ------QSNGGAINHLSDLSLDEFKNRFLMSAEAFEHLKTQFDLNAETNACSINGNAPAE 111

Qy        119 VDWREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNK 178
              :| |:   || :: || ||:|||| | | |:       : : |: | ||||:::
Db        112 IDLRQMRTVTPIRMQGGCGSCWAFSGVAATESAYLAYRNQSLDLAEQELVDCASQ----H 167

Qy        179 GCNGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVL 238
              ||:|  :    :||  | |:  :: | | | :  |:  :  |    | | ::     : :
Db        168 GCHGDTIPRGIEYIQHN-GVVQESYYRYVAREQSCRRPNAQRFG-ISNYCQIYPPNVNKI 225

Qy        239 KEAVA-NKGPVSVGVDARHPSFFLYRSGVY-------YEPSCTQNVNHGVLVVGYGDLNG 290
              :||:|     ::| :  :    | :  |         |:|:      | | :||| :  |
Db        226 REALAQTHSAIAVIIGIKDLDAFRHYDGRTIIQRDNGYQPN-----YHAVNIVGYSNAQG 280

Qy        291 KEYWLVKNSWGHNFGEEGYIRMARN 315
              :||:|:|||  |:|: ||   | |
Db        281 VDYWIVRNSWDTNWGDNGYGYFAAN 305


RESULT 11
US-10-878-556A-18
; Sequence 18, Application US/10878556A
; Publication No. US20050266399A1
; GENERAL INFORMATION:
;  APPLICANT: Hoffmann La-Roche Inc.
;  TITLE OF INVENTION: HCV regulated protein expression
;  FILE REFERENCE: 21762
;  CURRENT APPLICATION NUMBER: US/10/878,556A
;  CURRENT FILING DATE:  2004-06-28
;  NUMBER OF SEQ ID NOS: 199
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 18
;   LENGTH: 339
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;  PUBLICATION INFORMATION:
;   DATABASE ACCESSION NUMBER: sw_human/catb_human
;   DATABASE ENTRY DATE: 1988-08-01
US-10-878-556A-18

  Query Match             17.1%;  Score 309;  DB 6;  Length 339;
  Best Local Similarity   25.3%;  Pred. No. 1.6e-21;
  Matches   92;  Conservative   57;  Mismatches  123;  Indels   92;  Gaps   13;

Qy          5 VCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVMLHNL 64
              :| ||| ::| ::    |  |                                : : :|   |
Db          8 LCCLLVLANARSRPSFHPVSD----------------------------ELVNYVNKRNT 39

Qy         65 EHSMGMHSYDLGMNHL---------GDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRIL 115
                   | : |:: |::|         |    : ||     |::|: :
Db         40 TWQAGHNFYNVDMSYLKRLCGTFLGGPKPPQRVM-FTEDLKLPASFDAR---------- 87

Qy        116 PDSVDWREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVS--LSAQNLVDCSTE 173
                  | :   : |:: |||||:|||| || |:  :: : |   ||  :||::|: |
Db         88 ---EQWPQCPTIKEIRDQGSCGSCWAFGAVEAISDRICIHTNAHVSVEVSAEDLLTCCGS 144

Qy        174 KYGNKGCNGGFMTTAFQYIIDNKGIDSDASY-------PYKAMDLKCQYDSKYRAAT--- 223
                |: |||||:   |: :    ||: |   |       ||     :   :      |
Db        145 MCGD-GCNGGYPAEAWNF-WTRKGLVSGGLYESHVGCRPYSIPPCEHHVNGSRPPCTGEG 202

Qy        224 ----CSKYTELPY------------------GREDVLKEAVANKGPVSVGVDARHPSFF 260
                  |||  |  |                    :|:: |   | |||  |  : :  |
Db        203 DTPKCSKICEPGYSPTYKQDKHYGYNSYSVSNSEKDIMAEIYKN-GPVE-GAFSVYSDFL 260

Qy        261 LYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSWGHNFGEEGYIRMARNKGNHC 320
              ||:|||   :      | : ::|:|  ||  |||| |||  ::|: |: :: | : :||
Db        261 LYKSGVYQHVTGEMMGGHAIRILGWGVENGTPYWLVANSWNTDWGDNGFFKILRGQ-DHC 319

Qy        321 GIAS 324
              || |
Db        320 GIES 323


RESULT 12
US-10-995-561-681
; Sequence 681, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 681
;   LENGTH: 339
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-995-561-681
```

```
  Query Match            17.1%;  Score 308;  DB 6;  Length 339;
  Best Local Similarity   25.3%;  Pred. No. 2e-21;
  Matches   92;  Conservative   57;  Mismatches  123;  Indels   92;  Gaps   13;

Qy          5 VCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVMLHNL 64
              :| ||| ::| ::    | |                              : : :|   |
Db          8 LCCLLVLANARSRPSFHPLSD----------------------------ELVNYVNKRNT 39

Qy         65 EHSMGMHSYDLGMNHL---------GDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRIL 115
                 | : |:: |::|         |    : ||     |::|: :
Db         40 TWQAGHNFYNVDMSYLKRLCGTFLGGPKPPQRVM-FTEDLKLPASFDAR---------- 87

Qy        116 PDSVDWREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVS--LSAQNLVDCSTE 173
                | :   : |:: ||||||:|||| || |:  :: : |   ||  :||::|: |
Db         88 ---EQWPQCPTIKEIRDQGSCGSCWAFGAVEAISDRICIHTNAHVSVEVSAEDLLTCCGS 144

Qy        174 KYGNKGCNGGFMTTAFQYIIDNKGIDSDASY-------PYKAMDLKCQYDSKYRAAT--- 223
                |: |||||:   |: :    ||: |   |       ||     :   :       |
Db        145 MCGD-GCNGGYPAEAWNF-WTRKGLVSGGLYESHVGCRPYSIPPCEHHVNGSRPPCTGEG 202

Qy        224 ----CSKYTELPY------------------GREDVLKEAVANKGPVSVGVDARHPSFF 260
                   |||  | |                  :|:: |   | |||  |  : :  |
Db        203 DTPKCSKICEPGYSPTYKQDKHYGYNSYSVSNSEKDIMAEIYKN-GPVE-GAFSVYSDFL 260

Qy        261 LYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSWGHNFGEEGYIRMARNKGNHC 320
              ||:||||   :      | : ::|:|  ||  |||| |||  ::|: |: :: | : :||
Db        261 LYKSGVYQHVTGEMMGGHAIRILGWGVENGTPYWLVANSWNTDWGDNGFFKILRGQ-DHC 319

Qy        321 GIAS 324
              || |
Db        320 GIES 323


RESULT 13
US-10-995-561-682
; Sequence 682, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 682
;   LENGTH: 339
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-995-561-682

  Query Match            17.1%;  Score 308;  DB 6;  Length 339;
  Best Local Similarity   25.3%;  Pred. No. 2e-21;
  Matches   92;  Conservative   57;  Mismatches  123;  Indels   92;  Gaps   13;

Qy          5 VCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVMLHNL 64
              :| ||| ::| ::    | |                              : : :|   |
Db          8 LCCLLVLANARSRPSFHPLSD----------------------------ELVNYVNKRNT 39

Qy         65 EHSMGMHSYDLGMNHL---------GDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRIL 115
                 | : |:: |::|         |    : ||     |::|: :
Db         40 TWQAGHNFYNVDMSYLKRLCGTFLGGPKPPQRVM-FTEDLKLPASFDAR---------- 87

Qy        116 PDSVDWREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVS--LSAQNLVDCSTE 173
                | :   : |:: |||||:|||| || |:  :: : |   ||  :||::|: |
Db         88 ---EQWPQCPTIKEIRDQGSCGSCWAFGAVEAISDRICIHTNAHVSVEVSAEDLLTCCGS 144

Qy        174 KYGNKGCNGGFMTTAFQYIIDNKGIDSDASY-------PYKAMDLKCQYDSKYRAAT--- 223
                |: |||||:   |: :    ||: |   |       ||     :   :       |
Db        145 MCGD-GCNGGYPAEAWNF-WTRKGLVSGGLYESHVGCRPYSIPPCEHHVNGSRPPCTGEG 202

Qy        224 ----CSKYTELPY------------------GREDVLKEAVANKGPVSVGVDARHPSFF 260
                   |||  | |                  :|:: |   | |||  |  : :  |
Db        203 DTPKCSKICEPGYSPTYKQDKHYGYNSYSVSNSEKDIMAEIYKN-GPVE-GAFSVYSDFL 260

Qy        261 LYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSWGHNFGEEGYIRMARNKGNHC 320
              ||:||||   :      | : ::|:|  ||  |||| |||  ::|: |: :: | : :||
Db        261 LYKSGVYQHVTGEMMGGHAIRILGWGVENGTPYWLVANSWNTDWGDNGFFKILRGQ-DHC 319

Qy        321 GIAS 324
              || |
Db        320 GIES 323


RESULT 14
US-10-995-561-684
; Sequence 684, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 684
;   LENGTH: 339
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-995-561-684

  Query Match            17.1%;  Score 308;  DB 6;  Length 339;
  Best Local Similarity   25.3%;  Pred. No. 2e-21;
  Matches   92;  Conservative   57;  Mismatches  123;  Indels   92;  Gaps   13;

Qy          5 VCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVMLHNL 64
              :| ||| ::| ::    | |                              : : :|   |
Db          8 LCCLLVLANARSRPSFHPLSD----------------------------ELVNYVNKRNT 39

Qy         65 EHSMGMHSYDLGMNHL---------GDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRIL 115
                 | : |:: |::|         |    : ||     |::|: :
Db         40 TWQAGHNFYNVDMSYLKRLCGTFLGGPKPPQRVM-FTEDLKLPASFDAR---------- 87

Qy        116 PDSVDWREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVS--LSAQNLVDCSTE 173
                | :   : |:: |||||:|||| || |:  :: : |   ||  :||::|: |
Db         88 ---EQWPQCPTIKEIRDQGSCGSCWAFGAVEAISDRICIHTNAHVSVEVSAEDLLTCCGS 144

Qy        174 KYGNKGCNGGFMTTAFQYIIDNKGIDSDASY-------PYKAMDLKCQYDSKYRAAT--- 223
                |: |||||:   |: :    ||: |   |       ||     :   :       |
Db        145 MCGD-GCNGGYPAEAWNF-WTRKGLVSGGLYESHVGCRPYSIPPCEHHVNGSRPPCTGEG 202

Qy        224 ----CSKYTELPY------------------GREDVLKEAVANKGPVSVGVDARHPSFF 260
                   |||  | |                  :|:: |   | |||  |  : :  |
Db        203 DTPKCSKICEPGYSPTYKQDKHYGYNSYSVSNSEKDIMAEIYKN-GPVE-GAFSVYSDFL 260

Qy        261 LYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSWGHNFGEEGYIRMARNKGNHC 320
              ||:||||   :      | : ::|:|  ||  |||| |||  ::|: |: :: | : :||
Db        261 LYKSGVYQHVTGEMMGGHAIRILGWGVENGTPYWLVANSWNTDWGDNGFFKILRGQ-DHC 319

Qy        321 GIAS 324
              || |
Db        320 GIES 323


RESULT 15
US-10-995-561-685
; Sequence 685, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
```

```
;  TITLE OF INVENTION:  CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
;  TITLE OF INVENTION:  DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001559
;  CURRENT APPLICATION NUMBER: US/10/995,561
;  CURRENT FILING DATE:  2004-11-24
;  NUMBER OF SEQ ID NOS: 85702
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 685
;   LENGTH: 339
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-995-561-685

  Query Match             17.1%;  Score 308;  DB 6;  Length 339;
  Best Local Similarity   25.3%;  Pred. No. 2e-21;
  Matches   92;  Conservative   57;  Mismatches  123;  Indels   92;  Gaps   13;

Qy          5 VCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVMLHNL 64
              :| ||| ::| ::    |  |                             : : :|   |
Db          8 LCCLLVLANARSRPSFHPLSD----------------------------ELVNYVNKRNT 39

Qy         65 EHSMGMHSYDLGMNHL---------GDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRIL 115
                  | : |:: |::|          |    : ||     |::|: :
Db         40 TWQAGHNFYNVDMSYLKRLCGTFLGGPKPPQRVM-FTEDLKLPASFDAR---------- 87

Qy        116 PDSVDWREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVS--LSAQNLVDCSTE 173
                  | :   : |:: |||||:|||| || |:  :: : |   ||  :||::|: |
Db         88 ---EQWPQCPTIKEIRDQGSCGSCWAFGAVEAISDRICIHTNAHVSVEVSAEDLLTCCGS 144

Qy        174 KYGNKGCNGGFMTTAFQYIIDNKGIDSDASY-------PYKAMDLKCQYDSKYRAAT--- 223
                |: |||||:   |: :    ||: |   |       ||     :   :      |
Db        145 MCGD-GCNGGYPAEAWNF-WTRKGLVSGGLYESHVGCRPYSIPPCEHHVNGSRPPCTGEG 202

Qy        224 ----CSKYTELPY------------------GREDVLKEAVANKGPVSVGVDARHPSFF 260
                   |||  |  |                      :|:: |   | |||  |  : :  |
Db        203 DTPKCSKICEPGYSPTYKQDKHYGYNSYSVSNSEKDIMAEIYKN-GPVE-GAFSVYSDFL 260

Qy        261 LYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSWGHNFGEEGYIRMARNKGNHC 320
              ||:|||   :      | : ::|:|  ||  |||| |||  ::|: |: :: | : :||
Db        261 LYKSGVYQHVTGEMMGGHAIRILGWGVENGTPYWLVANSWNTDWGDNGFFKILRGQ-DHC 319

Qy        321 GIAS 324
              || |
Db        320 GIES 323


Search completed: January 10, 2006, 09:49:50
Job time : 51 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 10, 2006, 09:35:49 ; Search time 113 Seconds
(without alignments)
1223.907 Million cell updates/sec

Title: US-10-646-470-1
Perfect score: 1806
Sequence: 1 MKRLVCVLLVCSSAVAQLHK.....MARNKGNHCGIASFPSYPEI 331

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA_Main:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1806	100.0	331	4	US-10-099-275-2	Sequence 2, Appli
2	1800	99.7	331	4	US-10-273-577-6	Sequence 6, Appli
3	1800	99.7	331	5	US-10-973-858-20	Sequence 20, Appl
4	1800	99.7	331	5	US-10-287-436A-412	Sequence 412, App
5	1800	99.7	331	5	US-10-287-436A-1109	Sequence 1109, Ap
6	1797	99.5	331	5	US-10-635-398-96	Sequence 96, Appl
7	1794	99.3	331	4	US-10-273-577-1	Sequence 1, Appli
8	1792	99.2	331	5	US-10-809-816A-4	Sequence 4, Appli
9	1791	99.2	331	3	US-09-953-956-8	Sequence 8, Appli
10	1791	99.2	331	4	US-10-114-464-8	Sequence 8, Appli
11	1791	99.2	331	5	US-10-726-645-8	Sequence 8, Appli
12	1790	99.1	331	4	US-10-273-577-7	Sequence 7, Appli
13	1784	98.8	331	4	US-10-273-577-8	Sequence 8, Appli
14	1784	98.8	331	4	US-10-318-584-4	Sequence 4, Appli
15	1740	96.3	331	5	US-10-809-816A-2	Sequence 2, Appli
16	1700.5	94.2	330	4	US-10-010-577-2	Sequence 2, Appli
17	1700.5	94.2	330	5	US-10-894-104-2	Sequence 2, Appli
18	1700.5	94.2	330	5	US-10-809-816A-3	Sequence 3, Appli
19	1641	90.9	322	5	US-10-809-816A-7	Sequence 7, Appli
20	1599	88.5	331	4	US-10-010-580-2	Sequence 2, Appli
21	1599	88.5	331	5	US-10-894-106-2	Sequence 2, Appli
22	1599	88.5	331	5	US-10-809-816A-5	Sequence 5, Appli
23	1578	87.4	331	3	US-09-990-064-2	Sequence 2, Appli
24	1424	78.8	267	4	US-10-408-765A-1343	Sequence 1343, Ap
25	1335.5	73.9	330	5	US-10-809-816A-6	Sequence 6, Appli
26	1188	65.8	217	4	US-10-273-577-2	Sequence 2, Appli
27	1182	65.4	217	4	US-10-273-577-5	Sequence 5, Appli
28	1182	65.4	222	5	US-10-635-398-98	Sequence 98, Appl
29	1178	65.2	217	4	US-10-273-577-3	Sequence 3, Appli
30	1172	64.9	217	4	US-10-273-577-4	Sequence 4, Appli
31	952.5	52.7	329	3	US-09-953-956-7	Sequence 7, Appli
32	952.5	52.7	329	4	US-10-114-464-7	Sequence 7, Appli
33	952.5	52.7	329	4	US-10-318-584-3	Sequence 3, Appli
34	952.5	52.7	329	5	US-10-726-645-7	Sequence 7, Appli
35	952	52.7	330	5	US-10-929-919A-32	Sequence 32, Appl
36	951.5	52.7	329	3	US-09-953-956-2	Sequence 2, Appli
37	951.5	52.7	329	4	US-10-114-464-2	Sequence 2, Appli
38	951.5	52.7	329	4	US-10-318-584-2	Sequence 2, Appli
39	951.5	52.7	329	5	US-10-723-860-1211	Sequence 1211, Ap
40	951.5	52.7	329	5	US-10-726-645-2	Sequence 2, Appli
41	951.5	52.7	329	5	US-10-973-858-18	Sequence 18, Appl
42	951.5	52.7	329	5	US-10-756-149-5007	Sequence 5007, Ap
43	951.5	52.7	361	3	US-09-925-300-1693	Sequence 1693, Ap
44	868	48.1	291	5	US-10-836-897-4	Sequence 4, Appli
45	865	47.9	334	3	US-09-784-641-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-10-099-275-2
; Sequence 2, Application US/10099275
; Publication No. US20020187499A1
; GENERAL INFORMATION:
; APPLICANT: Schneider, Patrick
; APPLICANT: Yamamoto, Karen K.
; APPLICANT: French, Cynthia K.
; APPLICANT: Reprogen, Inc.
; TITLE OF INVENTION: Use of Cathepsin S in the Diagnosis and Treatment of
; TITLE OF INVENTION: Endometriosis
; FILE REFERENCE: 018002-001310US
; CURRENT APPLICATION NUMBER: US/10/099,275
; CURRENT FILING DATE: 2002-04-13
; PRIOR APPLICATION NUMBER: US/09/701,685
; PRIOR FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: US 60/088,017
; PRIOR FILING DATE: 1998-06-04
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-099-275-2

Query Match 100.0%; Score 1806; DB 4; Length 331;
Best Local Similarity 100.0%; Pred. No. 1.4e-166;
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
```

```
Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              ||||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331


RESULT 2
US-10-273-577-6
; Sequence 6, Application US/10273577
; Publication No. US20030143714A1
; GENERAL INFORMATION:
;  APPLICANT: Lamers, Marieke B.
;  APPLICANT:  Williams, David H.
;  TITLE OF INVENTION: Crystal Structure of a Mutant of Cathepsin S Enzyme
;  FILE REFERENCE: 1718-0202P
;  CURRENT APPLICATION NUMBER: US/10/273,577
;  CURRENT FILING DATE:  2003-02-14
;  PRIOR APPLICATION NUMBER: 60/330,191
;  PRIOR FILING DATE: 2001-10-19
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
;   LENGTH: 331
;   TYPE: PRT
;   ORGANISM: human
US-10-273-577-6

  Query Match             99.7%;  Score 1800;  DB 4;  Length 331;
  Best Local Similarity   99.7%;  Pred. No. 5.4e-166;
  Matches  330;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLTSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              ||||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331


RESULT 3
US-10-973-858-20
; Sequence 20, Application US/10973858
; Publication No. US20050176030A1
; GENERAL INFORMATION:
;  APPLICANT: Gan, Li
;  APPLICANT:  Gonzalez-Zulueta, Mirella
;  APPLICANT:  Ye, Shiming
;  APPLICANT:  Urfer, Roman
;  APPLICANT:  Nickolich, Karoly
;  TITLE OF INVENTION: Regulated Nucleic Acids in Pathogenesis of Alzheimer's Disease
;  FILE REFERENCE: AGYT-047
;  CURRENT APPLICATION NUMBER: US/10/973,858
;  CURRENT FILING DATE:  2004-10-25
;  PRIOR APPLICATION NUMBER: 60/515,562
;  PRIOR FILING DATE: 2003-10-28
;  NUMBER OF SEQ ID NOS: 48
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 20
;   LENGTH: 331
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-973-858-20

  Query Match             99.7%;  Score 1800;  DB 5;  Length 331;
  Best Local Similarity   99.7%;  Pred. No. 5.4e-166;
  Matches  330;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              ||||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331


RESULT 4
US-10-287-436A-412
; Sequence 412, Application US/10287436A
; Publication No. US20050202421A1
; GENERAL INFORMATION:
;  APPLICANT: CHILDREN'S HOSPITAL MEDICAL CENTER
;  TITLE OF INVENTION: METHOD FOR DIAGNOSIS AND TREATMENT OF
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS
;  FILE REFERENCE: 10872.514696
;  CURRENT APPLICATION NUMBER: US/10/287,436A
;  CURRENT FILING DATE:  2002-10-31
;  PRIOR APPLICATION NUMBER: US 60/336,220
;  PRIOR FILING DATE: 2001-10-31
;  NUMBER OF SEQ ID NOS: 1446
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 412
;   LENGTH: 331
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-287-436A-412

  Query Match             99.7%;  Score 1800;  DB 5;  Length 331;
  Best Local Similarity   99.7%;  Pred. No. 5.4e-166;
  Matches  330;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
```

```
                ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              |||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331


RESULT 5
US-10-287-436A-1109
; Sequence 1109, Application US/10287436A
; Publication No. US20050202421A1
; GENERAL INFORMATION:
;  APPLICANT: CHILDREN'S HOSPITAL MEDICAL CENTER
;  TITLE OF INVENTION: METHOD FOR DIAGNOSIS AND TREATMENT OF
;  TITLE OF INVENTION:  RHEUMATOID ARTHRITIS
;  FILE REFERENCE: 10872.514696
;  CURRENT APPLICATION NUMBER: US/10/287,436A
;  CURRENT FILING DATE:  2002-10-31
;  PRIOR APPLICATION NUMBER: US 60/336,220
;  PRIOR FILING DATE: 2001-10-31
;  NUMBER OF SEQ ID NOS: 1446
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1109
;   LENGTH: 331
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-287-436A-1109

  Query Match             99.7%;  Score 1800;  DB 5;  Length 331;
  Best Local Similarity   99.7%;  Pred. No. 5.4e-166;
  Matches  330;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              |||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331


RESULT 6
US-10-635-398-96
; Sequence 96, Application US/10635398
; Publication No. US20050037957A1
; GENERAL INFORMATION:
;  APPLICANT: David Anderson
;  APPLICANT:  Constance Berghs
;  APPLICANT:  Elina Catterton
;  APPLICANT:  Shlomit Edinger
;  APPLICANT:  Linda Gorman
;  APPLICANT:  Xiaojia (Sasha) Guo
;  APPLICANT:  John Herrmann
;  APPLICANT:  Ramesh Kekuda
;  APPLICANT:  Li Li
;  APPLICANT:  Daniel Rieger
;  APPLICANT:  Mei Zhong
;  TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
;  FILE REFERENCE: 21402-593 C
;  CURRENT APPLICATION NUMBER: US/10/635,398
;  CURRENT FILING DATE:  2003-08-06
;  PRIOR APPLICATION NUMBER: 10/160,619
;  PRIOR FILING DATE: 2002-06-03
;  PRIOR APPLICATION NUMBER: 60/295,661
;  PRIOR FILING DATE: 2001-06-04
;  PRIOR APPLICATION NUMBER: 60/359,122
;  PRIOR FILING DATE: 2002-02-21
;  PRIOR APPLICATION NUMBER: 60/296,404
;  PRIOR FILING DATE: 2001-06-06
;  PRIOR APPLICATION NUMBER: 60/359,035
;  PRIOR FILING DATE: 2002-02-22
;  PRIOR APPLICATION NUMBER: 60/299,949
;  PRIOR FILING DATE: 2002-06-21
;  PRIOR APPLICATION NUMBER: 60/295,607
;  PRIOR FILING DATE: 2001-06-04
;  PRIOR APPLICATION NUMBER: 60/359,964
;  PRIOR FILING DATE: 2002-02-27
;  PRIOR APPLICATION NUMBER: 60/296,418
;  PRIOR FILING DATE: 2001-06-06
;  PRIOR APPLICATION NUMBER: 60/341,562
;  PRIOR FILING DATE: 2001-12-14
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 146
;  SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 96
;   LENGTH: 331
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-635-398-96

  Query Match             99.5%;  Score 1797;  DB 5;  Length 331;
  Best Local Similarity   99.4%;  Pred. No. 1.1e-165;
  Matches  329;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANRGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              |||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331


RESULT 7
US-10-273-577-1
; Sequence 1, Application US/10273577
; Publication No. US20030143714A1
; GENERAL INFORMATION:
;  APPLICANT: Lamers, Marieke B.
```

```
;  APPLICANT:  Williams, David H.
;  TITLE OF INVENTION: Crystal Structure of a Mutant of Cathepsin S Enzyme
;  FILE REFERENCE: 1718-0202P
;  CURRENT APPLICATION NUMBER: US/10/273,577
;  CURRENT FILING DATE:  2003-02-14
;  PRIOR APPLICATION NUMBER: 60/330,191
;  PRIOR FILING DATE: 2001-10-19
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 331
;   TYPE: PRT
;   ORGANISM: human
US-10-273-577-1

  Query Match            99.3%;  Score 1794;  DB 4;  Length 331;
  Best Local Similarity   99.4%;  Pred. No. 2.1e-165;
  Matches  329;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLTSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              ||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331


RESULT 8
US-10-809-816A-4
; Sequence 4, Application US/10809816A
; Publication No. US20050214774A1
; GENERAL INFORMATION:
;  APPLICANT: LI, Shyr-Jiann et al.
;  TITLE OF INVENTION: ISOLATED MONKEY CATHEPSIN S PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING MONKEY CATHEPSIN S PROTEINS,
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: CL001507
;  CURRENT APPLICATION NUMBER: US/10/809,816A
;  CURRENT FILING DATE:  2004-03-26
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 331
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-809-816A-4

  Query Match            99.2%;  Score 1792;  DB 5;  Length 331;
  Best Local Similarity   99.4%;  Pred. No. 3.3e-165;
  Matches  329;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNWILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              ||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331


RESULT 9
US-09-953-956-8
; Sequence 8, Application US/09953956
; Patent No. US20020072107A1
; GENERAL INFORMATION:
;  APPLICANT: Hastings, et al.
;  TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
;  FILE REFERENCE: PF107D2D1
;  CURRENT APPLICATION NUMBER: US/09/953,956
;  CURRENT FILING DATE:  2001-09-18
;  PRIOR APPLICATION NUMBER: 09/219,441
;  PRIOR FILING DATE: 1998-12-23
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
;   LENGTH: 331
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-953-956-8

  Query Match            99.2%;  Score 1791;  DB 3;  Length 331;
  Best Local Similarity   99.1%;  Pred. No. 4.1e-165;
  Matches  328;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLTSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVTLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              ||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331


RESULT 10
US-10-114-464-8
; Sequence 8, Application US/10114464
; Publication No. US20020142448A1
```

```
; GENERAL INFORMATION:
;  APPLICANT: Hastings, et al.
;  TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
;  FILE REFERENCE: PF107D5
;  CURRENT APPLICATION NUMBER: US/10/114,464
;  CURRENT FILING DATE:  2002-04-03
;  PRIOR APPLICATION NUMBER: 08/553,125
;  PRIOR FILING DATE: 1995-11-07
;  PRIOR APPLICATION NUMBER: 08/208,007
;  PRIOR FILING DATE: 1994-03-08
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
;   LENGTH: 331
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-114-464-8

  Query Match             99.2%;  Score 1791;  DB 4;  Length 331;
  Best Local Similarity   99.1%;  Pred. No. 4.1e-165;
  Matches  328;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLTSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||||:||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVTLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              |||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331


RESULT 11
US-10-726-645-8
; Sequence 8, Application US/10726645
; Publication No. US20050019889A1
; GENERAL INFORMATION:
;  APPLICANT: Hastings, et al.
;  TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
;  FILE REFERENCE: PF107D5
;  CURRENT APPLICATION NUMBER: US/10/726,645
;  CURRENT FILING DATE:  2003-12-04
;  PRIOR APPLICATION NUMBER: US/10/114,464
;  PRIOR FILING DATE: 2002-04-03
;  PRIOR APPLICATION NUMBER: 08/553,125
;  PRIOR FILING DATE: 1995-11-07
;  PRIOR APPLICATION NUMBER: 08/208,007
;  PRIOR FILING DATE: 1994-03-08
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
;   LENGTH: 331
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-726-645-8

  Query Match             99.2%;  Score 1791;  DB 5;  Length 331;
  Best Local Similarity   99.1%;  Pred. No. 4.1e-165;
  Matches  328;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLTSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||||:||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVTLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              |||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331


RESULT 12
US-10-273-577-7
; Sequence 7, Application US/10273577
; Publication No. US20030143714A1
; GENERAL INFORMATION:
;  APPLICANT: Lamers, Marieke B.
;  APPLICANT:  Williams, David H.
;  TITLE OF INVENTION: Crystal Structure of a Mutant of Cathepsin S Enzyme
;  FILE REFERENCE: 1718-0202P
;  CURRENT APPLICATION NUMBER: US/10/273,577
;  CURRENT FILING DATE:  2003-02-14
;  PRIOR APPLICATION NUMBER: 60/330,191
;  PRIOR FILING DATE: 2001-10-19
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
;   LENGTH: 331
;   TYPE: PRT
;   ORGANISM: human
US-10-273-577-7

  Query Match             99.1%;  Score 1790;  DB 4;  Length 331;
  Best Local Similarity   99.4%;  Pred. No. 5.1e-165;
  Matches  329;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLTSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGASWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
```

```
                    ||||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331


RESULT 13
US-10-273-577-8
; Sequence 8, Application US/10273577
; Publication No. US20030143714A1
; GENERAL INFORMATION:
;  APPLICANT: Lamers, Marieke B.
;  APPLICANT:  Williams, David H.
;  TITLE OF INVENTION: Crystal Structure of a Mutant of Cathepsin S Enzyme
;  FILE REFERENCE: 1718-0202P
;  CURRENT APPLICATION NUMBER: US/10/273,577
;  CURRENT FILING DATE:  2003-02-14
;  PRIOR APPLICATION NUMBER: 60/330,191
;  PRIOR FILING DATE: 2001-10-19
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
;   LENGTH: 331
;   TYPE: PRT
;   ORGANISM: human
US-10-273-577-8

  Query Match             98.8%;  Score 1784;  DB 4;  Length 331;
  Best Local Similarity   99.1%;  Pred. No. 2e-164;
  Matches  328;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLTSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGASWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              |||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331


RESULT 14
US-10-318-584-4
; Sequence 4, Application US/10318584
; Publication No. US20030175937A1
;   GENERAL INFORMATION:
;        APPLICANT: Bromme, Dieter
;                   Okamoto, Kathleen
;        TITLE OF INVENTION: CATHEPSIN O2 PROTEASE
;        NUMBER OF SEQUENCES: 12
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
;             STREET: Four Embarcadero Center, Suite 3400
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States
;             ZIP: 94111-4187
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/318,584
;             FILING DATE: 13-Dec-2002
;             CLASSIFICATION: 435
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Silva, Robin M.
;             REGISTRATION NUMBER: 38,304
;             REFERENCE/DOCKET NUMBER: A-60261-1/DJB/RMS
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 781-1989
;             TELEFAX: (415) 398-3249
;             TELEX: 910 277299
;   INFORMATION FOR SEQ ID NO: 4:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 331 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: unknown
;             TOPOLOGY: unknown
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-10-318-584-4

  Query Match             98.8%;  Score 1784;  DB 4;  Length 331;
  Best Local Similarity   99.1%;  Pred. No. 2e-164;
  Matches  328;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||  |||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              |||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331


RESULT 15
US-10-809-816A-2
; Sequence 2, Application US/10809816A
; Publication No. US20050214774A1
; GENERAL INFORMATION:
;  APPLICANT: LI, Shyr-Jiann et al.
;  TITLE OF INVENTION: ISOLATED MONKEY CATHEPSIN S PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING MONKEY CATHEPSIN S PROTEINS,
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: CL001507
;  CURRENT APPLICATION NUMBER: US/10/809,816A
;  CURRENT FILING DATE:  2004-03-26
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 331
;   TYPE: PRT
;   ORGANISM: Cynomologous monkey
US-10-809-816A-2
```

```
  Query Match             96.3%;  Score 1740;  DB 5;  Length 331;
  Best Local Similarity   96.7%;  Pred. No. 3.7e-160;
  Matches  320;  Conservative    2;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKQLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              |||||||||||||||||||||||||||||||||||||||||||||||||| |:|||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNANQILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              |||||| |||||||| |||||||||||| | ||||||||||||||||||||||||||||
Db        181 NGGFMTRAFQYIIDNNGIDSDASYPYKATDQKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              |||||||||||||| |||||||||||||||||||||||||||| ||||||||||||||
Db        241 VVANKGPVSVGVDASHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGVLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              | |||||||||||||||||||||||||||||
Db        301 GRNFGEEGYIRMARNKGNHCGIASFPSYPEI 331


Search completed: January 10, 2006, 09:48:48
Job time : 114 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 10, 2006, 09:32:19 ; Search time 135 Seconds
(without alignments)
1077.292 Million cell updates/sec

Title: US-10-646-470-1
Perfect score: 1806
Sequence: 1 MKRLVCVLLVCSSAVAQLHK.....MARNKGNHCGIASFPSYPEI 331

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*
9: geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1806	100.0	331	3	AAY59634	Aay59634 Human Cat
2	1800	99.7	331	7	ADD35932	Add35932 Human cat
3	1800	99.7	331	7	ADP65168	Adp65168 Human cat
4	1800	99.7	331	9	ADY14944	Ady14944 PRO polyp
5	1800	99.7	331	9	ADY14946	Ady14946 PRO polyp
6	1800	99.7	331	9	AEC01587	Aec01587 Human cat
7	1797	99.5	331	7	ADD45416	Add45416 Human Pro
8	1797	99.5	331	7	ADE56902	Ade56902 Human Pro
9	1797	99.5	331	8	ADJ71695	Adj71695 Human NOV
10	1794	99.3	331	7	ADD35927	Add35927 Human cat
11	1791	99.2	331	9	ADY19784	Ady19784 PRO polyp
12	1790	99.1	331	7	ADD35933	Add35933 Human cat
13	1784	98.8	331	7	ADD35934	Add35934 Human cat
14	1700.5	94.2	330	6	AAO30451	Aao30451 Monkey ca
15	1678	92.9	313	8	ABM83324	Abm83324 Human dia
16	1678	92.9	313	8	ABM83323	Abm83323 Human dia
17	1599	88.5	331	7	ABR56160	Abr56160 Canine ca
18	1578	87.4	331	6	ABG73437	Abg73437 Human Cat
19	1424	78.8	267	7	ADJ69537	Adj69537 Human hea
20	1299.5	72.0	330	7	ADD45414	Add45414 Rat Prote
21	1299.5	72.0	330	7	ADE56900	Ade56900 Rat Prote
22	1188	65.8	217	7	ADD35928	Add35928 Human mat
23	1182	65.4	217	7	ADD35931	Add35931 Human mat
24	1182	65.4	222	8	ADJ71697	Adj71697 Human NOV
25	1179	65.3	217	9	ADY58885	Ady58885 Human cat
26	1178	65.2	217	7	ADD35929	Add35929 Human mat
27	1172	64.9	217	7	ADD35930	Add35930 Human mat
28	953.5	52.8	329	2	AAW77071	Aaw77071 Rat Cathe
29	952	52.7	330	9	ADZ11834	Adz11834 Canine ca
30	951.5	52.7	329	2	AAR82720	Aar82720 Human ost
31	951.5	52.7	329	2	AAR95599	Aar95599 Prepro-ca
32	951.5	52.7	329	2	AAW41645	Aaw41645 Human cat
33	951.5	52.7	329	2	AAW39216	Aaw39216 Human cat
34	951.5	52.7	329	3	AAY81000	Aay81000 Human pro
35	951.5	52.7	329	7	ADE60397	Ade60397 Human Pro
36	951.5	52.7	329	7	ADD45861	Add45861 Human Pro
37	951.5	52.7	329	8	ADP12997	Adp12997 Protein e
38	951.5	52.7	329	8	ADQ18392	Adq18392 Human sof
39	951.5	52.7	329	8	ABM80141	Abm80141 Tumour-as
40	951.5	52.7	329	8	ADP25400	Adp25400 PRO polyp
41	951.5	52.7	329	9	AEC01585	Aec01585 Human cat
42	951.5	52.7	361	3	AAB57115	Aab57115 Human pro
43	951	52.7	329	7	ADE60395	Ade60395 Rat Prote
44	951	52.7	329	7	ADD45859	Add45859 Rat Prote
45	948.5	52.5	329	3	AAB30506	Aab30506 A monkey

ALIGNMENTS

RESULT 1
AAY59634
ID AAY59634 standard; protein; 331 AA.
XX
AC AAY59634;
XX
DT 27-MAR-2000 (first entry)
XX
DE Human Cathepsin S amino acid sequence.
XX
KW Cathepsin S; human; endometriosis; treatment; diagnose.
XX
OS Homo sapiens.
XX
PN WO9963115-A2.
XX
PD 09-DEC-1999.
XX
PF 03-JUN-1999; 99WO-US012335.
XX
PR 04-JUN-1998; 98US-0088017P.
XX
PA (REPR-) REPROGEN INC.
XX
PI Schneider P, Yamamoto KK, French CK;
XX
DR WPI; 2000-086986/07.
DR N-PSDB; AAZ56150.
XX
PT Use of cathepsin S in the diagnosis and treatment of endometriosis.
XX
PS Example; Page 15; 60pp; English.
XX
CC This is the human cathepsin S protein sequence. Detecting levels of the
CC cathepsin S gene product in a sample compared to a control sample can be
CC used as a method of diagnosing endometriosis. The invention also relates
CC to a method for treating endometriosis through the administration of a
CC probe comprising a detectable label and a ligand that specifically binds
CC a cathepsin S gene product to the subject. The endometriotic lesion can
CC be identified in situ by locating bound labelled probe; and the lesion
CC can be excised. The methods are useful to diagnose, monitor the progress
CC of and treat endometriosis in a subject. The methods are also useful for
CC screening for modulators of cathepsin S gene production in endometrial
CC cells. Antisense cathepsin S gene oligonucleotides are useful for the
CC treatment of endometriosis by down-regulating cathepsin S genes
XX
SQ Sequence 331 AA;

```
  Query Match             100.0%;  Score 1806;  DB 3;  Length 331;
  Best Local Similarity   100.0%;  Pred. No. 3e-171;
  Matches  331;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              ||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
```

RESULT 2
ADD35932
ID ADD35932 standard; peptide; 331 AA.
XX
AC ADD35932;
XX
DT 15-JAN-2004 (first entry)
XX
DE Human cathepsin S mutant #1.
XX
KW Crystal; cathepsin S; catS inhibitor; human; mutant; mutein.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN US2003143714-A1.
XX
PD 31-JUL-2003.
XX
PF 18-OCT-2002; 2002US-00273577.
XX
PR 19-OCT-2001; 2001US-0330191P.
XX
PA (MEDI-) MEDIVIR UK LTD.
XX
PI Lamers MBAC, Williams DH, Turkenburg JP, Hubbard RE;
XX
DR WPI; 2003-829792/77.
XX
PT Crystalline cathepsin S polypeptide free of irreversible inhibitor useful
PT for identifying and producing potential catS inhibitor.
XX
PS Claim 2; SEQ ID NO 6; 59pp; English.
XX
CC The invention relates to a crystalline cathepsin S polypeptide. The
CC crystalline catS polypeptide is useful for identifying a potential catS
CC inhibitor molecule. The present sequence represents the amino acid
CC sequence of a human cathepsin S mutant.
XX
SQ Sequence 331 AA;

```
  Query Match             99.7%;  Score 1800;  DB 7;  Length 331;
  Best Local Similarity   99.7%;  Pred. No. 1.2e-170;
  Matches  330;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLTSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              ||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
```

RESULT 3
ADP65168
ID ADP65168 standard; protein; 331 AA.
XX
AC ADP65168;
XX
DT 12-AUG-2004 (first entry)
XX
DE Human cathepsin S preproprotein.
XX
KW autoimmune disease; arthritide; gene expression analysis;
KW rheumatoid arthritis; collagen-induced; immunosuppressive; antirheumatic;
KW antiarthritic; osteopathic; antigout; antiinflammatory; dermatological;
KW immunomodulatory; lupus; ankylosing spondylitis; Fibrositis;
KW fibromyalgia; osteoarthritis; gout; juvenile rheumatoid arthritis;
KW immune; human.
XX
OS Homo sapiens.
XX
PN WO2003072827-A1.
XX
PD 04-SEP-2003.
XX
PF 31-OCT-2002; 2002WO-US035433.
XX
PR 31-OCT-2001; 2001US-0336220P.
XX
PA (CHIL-) CHILDREN'S HOSPITAL MEDICAL CENT.
XX
PI Hirsch R, Thorton SL;
XX
DR WPI; 2003-712740/67.
DR GENBANK; NP_004070.
XX
PT Diagnosing and analyzing autoimmune disease using gene expression
PT profiles and microarray technology, useful for diagnosing and treating
PT rheumatoid arthritis, lupus, fibrositis, osteoarthritis, fibromyalgia and
PT gout.
XX
PS Disclosure; Page; 56pp; English.
XX
CC The invention relates to a novel method for diagnosing and analysing
CC autoimmune disease or arthritides. The method comprises obtaining a
CC patient sample containing mRNA, analysing gene expression using the mRNA
CC that results in a gene expression signature of the mRNA, and using that
CC gene expression signature to diagnose or analyse the autoimmune disease

CC or arthritides in the patient, where gene expression of at least 60% of
CC the genes correlates with that of the gene signature. The invention
CC further comprises: a treatment of rheumatoid arthritis; identification of
CC genes for targeting in the treatment of rheumatoid arthritis in a mammal
CC other than a mouse; diagnosis of rheumatoid arthritis in a mammal; an
CC array or gene chip, specific for rheumatoid arthritis; diagnosis or
CC analyses of autoimmune disease or rheumatoid arthritis; screening the
CC efficacy of a candidate drug in vitro for the treatment of collagen-
CC induced arthritis; and reducing the symptoms associated with collagen-
CC induced arthritis. The compositions of the invention have the following
CC activities: immunosuppressive, antirheumatic, antiarthritic, osteopathic,
CC antigout, antiinflammatory, dermatological, and immunomodulatory. The
CC methods and compositions of the present invention are useful for
CC diagnosing and treating autoimmune disease or arthritides, such as
CC rheumatoid arthritis, lupus, ankylosing spondylitis, fibrositis,
CC fibromyalgia, osteoarthritis, gout, juvenile rheumatoid arthritis, and an
CC immune disease caused by an infectious agent. This sequence represents a
CC protein sequence relating to the genes used in the analysis and treatment
CC of autoimmune diseases or arthritides. Note: This sequence is not shown
CC in the specification. It has been supplied in an electronic format from
CC WIPO.
XX
SQ Sequence 331 AA;

```
  Query Match             99.7%;  Score 1800;  DB 7;  Length 331;
  Best Local Similarity   99.7%;  Pred. No. 1.2e-170;
  Matches  330;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              |||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
```

RESULT 4
ADY14944
ID ADY14944 standard; protein; 331 AA.
XX
AC ADY14944;
XX
DT 05-MAY-2005 (first entry)
XX
DE PRO polypeptide SEQ ID NO 750.
XX
KW Antiinflammatory; Immune disorder; Dermatological; Immunosuppressive;
KW Antirheumatic; Antiarthritic; Osteopathic; Hemostatic; Antianemic;
KW Antithyroid; Antidiabetic; Nephrotropic; CNS-Gen.; Hepatotropic;
KW Virucide; Gastrointestinal-Gen.; Antipsoriatic; Antiasthmatic;
KW Antiallergic; ds; gene; diagnosis.
XX
OS Homo sapiens.
XX
PN WO2005016962-A2.
XX
PD 24-FEB-2005.
XX
PF 11-AUG-2004; 2004WO-US026249.
XX
PR 11-AUG-2003; 2003US-0493546P.
XX
PA (GETH) GENENTECH INC.
XX
PI Abbas A, Clark H, Ouyang W, Williams MP, Wood WI, Wu TD;
XX
DR WPI; 2005-182330/19.
XX
PT New nucleic acid encoding PRO polypeptide, useful for diagnosing and
PT treating an immune related disorder, e.g. systemic lupus erythematosus,
PT rheumatoid arthritis, osteoarthritis, thyroiditis, or diabetes mellitus.
XX
PS Claim 8; SEQ ID NO 750; 158pp; English.
XX
CC The invention relates to an isolated nucleic acid encoding a PRO
CC polypeptide. The polypeptide, agonist or an antagonist, antibody,
CC composition, and method are useful for diagnosing and treating an immune
CC related disorder, e.g. systemic lupus erythematosus, rheumatoid
CC arthritis. The present sequence represents a DNA encoding a PRO
CC polypeptide.
XX
SQ Sequence 331 AA;

```
  Query Match             99.7%;  Score 1800;  DB 9;  Length 331;
  Best Local Similarity   99.7%;  Pred. No. 1.2e-170;
  Matches  330;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              |||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
```

RESULT 5
ADY14946
ID ADY14946 standard; protein; 331 AA.
XX
AC ADY14946;
XX
DT 05-MAY-2005 (first entry)
XX
DE PRO polypeptide SEQ ID NO 752.
XX
KW Antiinflammatory; Immune disorder; Dermatological; Immunosuppressive;
KW Antirheumatic; Antiarthritic; Osteopathic; Hemostatic; Antianemic;
KW Antithyroid; Antidiabetic; Nephrotropic; CNS-Gen.; Hepatotropic;
KW Virucide; Gastrointestinal-Gen.; Antipsoriatic; Antiasthmatic;
KW Antiallergic; ds; gene; diagnosis.
XX
OS Homo sapiens.

XX
PN WO2005016962-A2.
XX
PD 24-FEB-2005.
XX
PF 11-AUG-2004; 2004WO-US026249.
XX
PR 11-AUG-2003; 2003US-0493546P.
XX
PA (GETH) GENENTECH INC.
XX
PI Abbas A, Clark H, Ouyang W, Williams MP, Wood WI, Wu TD;
XX
DR WPI; 2005-182330/19.
XX
PT New nucleic acid encoding PRO polypeptide, useful for diagnosing and
PT treating an immune related disorder, e.g. systemic lupus erythematosus,
PT rheumatoid arthritis, osteoarthritis, thyroiditis, or diabetes mellitus.
XX
PS Claim 8; SEQ ID NO 752; 158pp; English.
XX
CC The invention relates to an isolated nucleic acid encoding a PRO
CC polypeptide. The polypeptide, agonist or an antagonist, antibody,
CC composition, and method are useful for diagnosing and treating an immune
CC related disorder, e.g. systemic lupus erythematosus, rheumatoid
CC arthritis. The present sequence represents a DNA encoding a PRO
CC polypeptide.
XX
SQ Sequence 331 AA;

Query Match 99.7%; Score 1800; DB 9; Length 331;
Best Local Similarity 99.7%; Pred. No. 1.2e-170;
Matches 330; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              |||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
```

RESULT 6
AEC01587
ID AEC01587 standard; protein; 331 AA.
XX
AC AEC01587;
XX
DT 20-OCT-2005 (first entry)
XX
DE Human cathepsin S (CTSS) protein, SEQ ID NO: 20.
XX
KW Alzheimers disease; neuroprotective; nootropic; degeneration;
KW neurological disease; neurodegenerative disease; diagnosis; gene therapy;
KW prognosis; cathepsin S.
XX
OS Homo sapiens.
XX
PN US2005176030-A1.
XX
PD 11-AUG-2005.
XX
PF 25-OCT-2004; 2004US-00973858.
XX
PR 28-OCT-2003; 2003US-0515562P.
XX
PA (GANL/) GAN L.
PA (GONZ/) GONZALEZ-ZULUETA M.
PA (YESS/) YE S.
PA (URFE/) URFER R.
PA (NIKO/) NIKOLICH K.
XX
PI Gan L, Gonzalez-Zulueta M, Ye S, Urfer R, Nikolich K;
XX
DR WPI; 2005-581188/59.
DR N-PSDB; AEC01586.
DR REFSEQ; NP_004070.
XX
PT Detecting neurodegenerative disorder or its susceptibility, involves
PT detecting presence of differential expression of gene encoding
PT polypeptide having linear peptide sequence in biological sample.
XX
PS Claim 1; SEQ ID NO 20; 148pp; English.
XX
CC The present invention relates to a method of detecting a
CC neurodegenerative disorder (preferably Alzheimer's disease). The method
CC involves detecting the presence of differential expression of a gene
CC encoding a polypeptide having a linear peptide sequence in biological
CC sample. The invention is also useful in gene therapy. The present
CC sequence is the human cathepsin S (CTSS) protein. This protein encoding
CC gene is one of the Alzheimer's disease associated gene.
XX
SQ Sequence 331 AA;

Query Match 99.7%; Score 1800; DB 9; Length 331;
Best Local Similarity 99.7%; Pred. No. 1.2e-170;
Matches 330; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              |||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
```

RESULT 7
ADD45416
ID ADD45416 standard; protein; 331 AA.
XX
AC ADD45416;
XX

```
DT   29-JAN-2004  (first entry)
XX
DE   Human Protein P25774, SEQ ID NO 10849.
XX
KW   Human; pain; neuronal tissue; gene therapy;
KW   spinal segmental nerve injury; chronic constriction injury; CCI;
KW   spared nerve injury; SNI; Chung.
XX
OS   Homo sapiens.
XX
PN   WO2003016475-A2.
XX
PD   27-FEB-2003.
XX
PF   14-AUG-2002; 2002WO-US025765.
XX
PR   14-AUG-2001; 2001US-0312147P.
PR   01-NOV-2001; 2001US-0346382P.
PR   26-NOV-2001; 2001US-0333347P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
PA   (FARB ) BAYER AG.
XX
PI   Woolf C,  D'urso D,  Befort K,  Costigan M;
XX
DR   WPI; 2003-268312/26.
DR   GENBANK; P25774.
XX
PT   New composition comprising two or more isolated polypeptides, useful for
PT   preparing a medicament for treating pain in an animal.
XX
PS   Claim 1; Page; 1017pp; English.
XX
CC   The invention discloses a composition comprising two or more isolated rat
CC   or human polynucleotides or a polynucleotide which represents a fragment,
CC   derivative or allelic variation of the nucleic acid sequence. Also
CC   claimed are a vector comprising the novel polynucleotide, a host cell
CC   comprising the vector, a method for identifying a nucleotide sequence
CC   which is differentially regulated in an animal subjected to pain and a
CC   kit to perform the method, an array, a method for identifying an agent
CC   that increases or decreases the expression of the polynucleotide sequence
CC   that is differentially expressed in neuronal tissue of a first animal
CC   subjected to pain, a method for identifying a compound which regulates
CC   the expression of a polynucleotide sequence which is differentially
CC   expressed in an animal subjected to pain, a method for identifying a
CC   compound that regulates the activity of one or more of the
CC   polynucleotides, a method for producing a pharmaceutical composition, a
CC   method for identifying a compound or small molecule that regulates the
CC   activity in an animal of one or more of the polypeptides given in the
CC   specification, a method for identifying a compound useful in treating
CC   pain and a pharmaceutical composition comprising the one or more
CC   polypeptides or their antibodies. The polynucleotide or the compound that
CC   modulates its activity is useful for preparing a medicament for treating
CC   pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC   injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC   therapy). The sequence presented is a human protein (shown in Table 2 of
CC   the specification) which is differentially expressed during pain. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic form directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 331 AA;

  Query Match             99.5%;  Score 1797;  DB 7;  Length 331;
  Best Local Similarity   99.4%;  Pred. No. 2.4e-170;
  Matches  329;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||||:||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVTLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              |||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331


RESULT 8
ADE56902
ID   ADE56902 standard; protein; 331 AA.
XX
AC   ADE56902;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Human Protein P25774, SEQ ID NO 2757.
XX
KW   Human; pain; neuronal tissue; gene therapy;
KW   spinal segmental nerve injury; chronic constriction injury; CCI;
KW   spared nerve injury; SNI; Chung.
XX
OS   Homo sapiens.
XX
PN   WO2003016475-A2.
XX
PD   27-FEB-2003.
XX
PF   14-AUG-2002; 2002WO-US025765.
XX
PR   14-AUG-2001; 2001US-0312147P.
PR   01-NOV-2001; 2001US-0346382P.
PR   26-NOV-2001; 2001US-0333347P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
PA   (FARB ) BAYER AG.
XX
PI   Woolf C,  D'urso D,  Befort K,  Costigan M;
XX
DR   WPI; 2003-268312/26.
DR   GENBANK; P25774.
XX
PT   New composition comprising two or more isolated polypeptides, useful for
PT   preparing a medicament for treating pain in an animal.
XX
PS   Claim 1; Page; 1017pp; English.
XX
CC   The invention discloses a composition comprising two or more isolated rat
CC   or human polynucleotides or a polynucleotide which represents a fragment,
CC   derivative or allelic variation of the nucleic acid sequence. Also
CC   claimed are a vector comprising the novel polynucleotide, a host cell
CC   comprising the vector, a method for identifying a nucleotide sequence
CC   which is differentially regulated in an animal subjected to pain and a
CC   kit to perform the method, an array, a method for identifying an agent
CC   that increases or decreases the expression of the polynucleotide sequence
CC   that is differentially expressed in neuronal tissue of a first animal
CC   subjected to pain, a method for identifying a compound which regulates
CC   the expression of a polynucleotide sequence which is differentially
CC   expressed in an animal subjected to pain, a method for identifying a
CC   compound that regulates the activity of one or more of the
CC   polynucleotides, a method for producing a pharmaceutical composition, a
CC   method for identifying a compound or small molecule that regulates the
```

```
CC   activity in an animal of one or more of the polypeptides given in the
CC   specification, a method for identifying a compound useful in treating
CC   pain and a pharmaceutical composition comprising the one or more
CC   polypeptides or their antibodies. The polynucleotide or the compound that
CC   modulates its activity is useful for preparing a medicament for treating
CC   pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC   injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC   therapy). The sequence presented is a human protein (shown in Table 2 of
CC   the specification) which is differentially expressed during pain. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic form directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 331 AA;

  Query Match             99.5%;  Score 1797;  DB 7;  Length 331;
  Best Local Similarity   99.4%;  Pred. No. 2.4e-170;
  Matches  329;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||:||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVTLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              |||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331


RESULT 9
ADJ71695
ID   ADJ71695 standard; protein; 331 AA.
XX
AC   ADJ71695;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Human NOV9a protein SEQ ID NO:96.
XX
KW   human; cytostatic; antidiabetic; anorectic; CNS; cardiovascular;
KW   antiinflammatory; gene therapy; antisense therapy; cancer; diabetes;
KW   obesity; endocrine disorder; inflammatory disorder.
XX
OS   Homo sapiens.
XX
PN   WO2004015076-A2.
XX
PD   19-FEB-2004.
XX
PF   07-AUG-2003; 2003WO-US024788.
XX
PR   07-AUG-2002; 2002US-0401597P.
PR   09-AUG-2002; 2002US-0402248P.
PR   12-AUG-2002; 2002US-0402815P.
PR   13-AUG-2002; 2002US-0403485P.
PR   14-AUG-2002; 2002US-0403574P.
PR   15-AUG-2002; 2002US-0403732P.
PR   20-AUG-2002; 2002US-0404829P.
PR   27-AUG-2002; 2002US-0406392P.
PR   06-AUG-2003; 2003US-00406392.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Anderson DW,  Berghs C,  Catterton E,  Edinger SR,  Gorman L,  Guo X;
PI   Herrmann JL,  Kekuda R,  Li L,  Rieger DK,  Zhong M;
XX
DR   WPI; 2004-180659/17.
DR   N-PSDB; ADJ71694.
XX
PT   Novel polypeptides (NOVX) and nucleic acid molecules useful for treating,
PT   preventing and diagnosing pathological conditions with NOVX-associated
PT   disorders, such as cancer, obesity, diabetes and inflammatory diseases.
XX
PS   Claim 2; SEQ ID NO 96; 267pp; English.
XX
CC   The invention relates to a novel isolated NOVX polypeptide. A polypeptide
CC   of the invention has cytostatic, antidiabetic, anorectic, CNS-gen.,
CC   cardiovascular-gen., and antiinflammatory activity. A polynucleotide
CC   encoding a polypeptide of the invention may have a use in gene therapy,
CC   and antisense therapy. The methods and compositions of the present
CC   invention are useful for the diagnosis and treatment of disorders
CC   associated with aberrant expression or activity of the NOVX polypeptide,
CC   such as cancer, diabetes, obesity, and endocrine, CNS, cardiovascular and
CC   inflammatory disorders. They can also be used in various detection and
CC   screening assays, chromosome mapping, tissue typing and predictive
CC   medicine. The present sequence represents a NOVX polypeptide of the
CC   invention.
XX
SQ   Sequence 331 AA;

  Query Match             99.5%;  Score 1797;  DB 8;  Length 331;
  Best Local Similarity   99.4%;  Pred. No. 2.4e-170;
  Matches  329;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||:||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANRGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              |||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331


RESULT 10
ADD35927
ID   ADD35927 standard; peptide; 331 AA.
XX
AC   ADD35927;
XX
DT   15-JAN-2004  (first entry)
XX
DE   Human cathepsin S wild-tvpe.
XX
KW   Crystal; cathepsin S; catS inhibitor; human.
XX
```

OS Homo sapiens.
XX
PN US2003143714-A1.
XX
PD 31-JUL-2003.
XX
PF 18-OCT-2002; 2002US-00273577.
XX
PR 19-OCT-2001; 2001US-0330191P.
XX
PA (MEDI-) MEDIVIR UK LTD.
XX
PI Lamers MBAC, Williams DH, Turkenburg JP, Hubbard RE;
XX
DR WPI; 2003-829792/77.
XX
PT Crystalline cathepsin S polypeptide free of irreversible inhibitor useful
PT for identifying and producing potential catS inhibitor.
XX
PS Claim 2; SEQ ID NO 1; 59pp; English.
XX
CC The invention relates to a crystalline cathepsin S polypeptide. The
CC crystalline catS polypeptide is useful for identifying a potential catS
CC inhibitor molecule. The present sequence represents the amino acid
CC sequence of human cathepsin S wild-tvpe
XX
SQ Sequence 331 AA;

Query Match 99.3%; Score 1794; DB 7; Length 331;
Best Local Similarity 99.4%; Pred. No. 4.8e-170;
Matches 329; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLTSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              |||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
```

RESULT 11
ADY19784
ID ADY19784 standard; protein; 331 AA.
XX
AC ADY19784;
XX
DT 05-MAY-2005 (first entry)
XX
DE PRO polypeptide SEQ ID NO 5590.
XX
KW Antiinflammatory; Immune disorder; Dermatological; Immunosuppressive;
KW Antirheumatic; Antiarthritic; Osteopathic; Hemostatic; Antianemic;
KW Antithyroid; Antidiabetic; Nephrotropic; CNS-Gen.; Hepatotropic;
KW Virucide; Gastrointestinal-Gen.; Antipsoriatic; Antiasthmatic;
KW Antiallergic; diagnosis.
XX

OS Homo sapiens.
XX
PN WO2005016962-A2.
XX
PD 24-FEB-2005.
XX
PF 11-AUG-2004; 2004WO-US026249.
XX
PR 11-AUG-2003; 2003US-0493546P.
XX
PA (GETH) GENENTECH INC.
XX
PI Abbas A, Clark H, Ouyang W, Williams MP, Wood WI, Wu TD;
XX
DR WPI; 2005-182330/19.
XX
PT New nucleic acid encoding PRO polypeptide, useful for diagnosing and
PT treating an immune related disorder, e.g. systemic lupus erythematosus,
PT rheumatoid arthritis, osteoarthritis, thyroiditis, or diabetes mellitus.
XX
PS Claim 8; SEQ ID NO 5590; 158pp; English.
XX
CC The invention relates to an isolated nucleic acid encoding a PRO
CC polypeptide. The polypeptide, agonist or an antagonist, antibody,
CC composition, and method are useful for diagnosing and treating an immune
CC related disorder, e.g. systemic lupus erythematosus, rheumatoid
CC arthritis. The present sequence represents a PRO polypeptide.
XX
SQ Sequence 331 AA;

Query Match 99.2%; Score 1791; DB 9; Length 331;
Best Local Similarity 99.1%; Pred. No. 9.5e-170;
Matches 328; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLTSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVTLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              |||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
```

RESULT 12
ADD35933
ID ADD35933 standard; peptide; 331 AA.
XX
AC ADD35933;
XX
DT 15-JAN-2004 (first entry)
XX
DE Human cathepsin S mutant #2.
XX
KW Crystal; cathepsin S; catS inhibitor; human; mutant; mutein.
XX
OS Synthetic.
OS Homo sapiens.

XX
PN US2003143714-A1.
XX
PD 31-JUL-2003.
XX
PF 18-OCT-2002; 2002US-00273577.
XX
PR 19-OCT-2001; 2001US-0330191P.
XX
PA (MEDI-) MEDIVIR UK LTD.
XX
PI Lamers MBAC, Williams DH, Turkenburg JP, Hubbard RE;
XX
DR WPI; 2003-829792/77.
XX
PT Crystalline cathepsin S polypeptide free of irreversible inhibitor useful
PT for identifying and producing potential catS inhibitor.
XX
PS Claim 2; SEQ ID NO 7; 59pp; English.
XX
CC The invention relates to a crystalline cathepsin S polypeptide. The
CC crystalline catS polypeptide is useful for identifying a potential catS
CC inhibitor molecule. The present sequence represents the amino acid
CC sequence of a human cathepsin S mutant.
XX
SQ Sequence 331 AA;

Query Match 99.1%; Score 1790; DB 7; Length 331;
Best Local Similarity 99.4%; Pred. No. 1.2e-169;
Matches 329; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
Db 1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy 61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
Db 61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLTSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy 121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
Db 121 WREKGCVTEVKYQGSCGASWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180

Qy 181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
Db 181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy 241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
Db 241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy 301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
Db 301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331

RESULT 13
ADD35934
ID ADD35934 standard; peptide; 331 AA.
XX
AC ADD35934;
XX
DT 15-JAN-2004 (first entry)
XX
DE Human cathepsin S mutant #3.
XX
KW Crystal; cathepsin S; catS inhibitor; human; mutant; mutein.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN US2003143714-A1.
XX
PD 31-JUL-2003.
XX
PF 18-OCT-2002; 2002US-00273577.
XX
PR 19-OCT-2001; 2001US-0330191P.
XX
PA (MEDI-) MEDIVIR UK LTD.
XX
PI Lamers MBAC, Williams DH, Turkenburg JP, Hubbard RE;
XX
DR WPI; 2003-829792/77.
XX
PT Crystalline cathepsin S polypeptide free of irreversible inhibitor useful
PT for identifying and producing potential catS inhibitor.
XX
PS Claim 2; SEQ ID NO 8; 59pp; English.
XX
CC The invention relates to a crystalline cathepsin S polypeptide. The
CC crystalline catS polypeptide is useful for identifying a potential catS
CC inhibitor molecule. The present sequence represents the amino acid
CC sequence of a human cathepsin S mutant.
XX
SQ Sequence 331 AA;

Query Match 98.8%; Score 1784; DB 7; Length 331;
Best Local Similarity 99.1%; Pred. No. 4.8e-169;
Matches 328; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
Db 1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy 61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
Db 61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLTSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy 121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
Db 121 WREKGCVTEVKYQGSCGASWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180

Qy 181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
Db 181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy 241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
Db 241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy 301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
Db 301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331

RESULT 14
AAO30451
ID AAO30451 standard; protein; 330 AA.
XX
AC AAO30451;
XX
DT 22-SEP-2003 (first entry)
XX
DE Monkey cathepsin S protein.
XX
KW Monkey; cathepsin S; immune disease; lupus; rheumatoid arthritis; asthma;
KW gene therapy.
XX
OS Macaca sp.
XX
PN WO2003040396-A2.
XX
PD 15-MAY-2003.
XX
PF 25-OCT-2002; 2002WO-US034383.

XX
PR 08-NOV-2001; 2001US-00010577.
XX
PA (ORTH) ORTHO-MCNEIL PHARM INC.
XX
PI Thurmond R, Karlsson L, Baker S;
XX
DR WPI; 2003-430675/40.
DR N-PSDB; AAL61281.
XX
PT New nucleic acid molecule encoding a monkey cathepsin S protein, useful
PT for identifying modulators of cathepsin S for treating chronic immune
PT diseases such as lupus, rheumatoid arthritis, or asthma.
XX
PS Claim 6; Fig 2; 57pp; English.
XX
CC The invention relates to monkey cathepsin S proteins and nucleic acid
CC molecules encoding such proteins. Polypeptides of the invention are
CC useful for identifying cathepsin S modulators which are useful as
CC therapeutic agents for treating conditions mediated by cathepsin S e.g.
CC chronic immune diseases such as lupus, rheumatoid arthritis or asthma.
CC The invention is also useful in gene therapy. The present sequence is
CC monkey cathepsin S protein
XX
SQ Sequence 330 AA;

Query Match 94.2%; Score 1700.5; DB 6; Length 330;
Best Local Similarity 93.7%; Pred. No. 1e-160;
Matches 310; Conservative 11; Mismatches 9; Indels 1; Gaps 1;

```
Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||:|||||  ||||||  ||||||||||||:|||||||||||||||||||||||||||||
Db          1 MKQLVCVLFVCSSAVTQLHKDPTLDHHWNLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              ||||||||||||||||||||||||||||||||||||||:|||||||||||||::||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPNQWQRNITYKSNPNQMLPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCS-EKYGNKGC 179

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              ||||||  |||||||||||||:|||||| | ||||||||||||||||||||||||||||
Db        180 NGGFMTEAFQYIIDNKGIDSEASYPYKATDQKCQYDSKYRAATCSKYTELPYGREDVLKE 239

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              |||||||| ||||| ||||||||||||:|:|||  |||||||:||||||||||||||||
Db        240 AVANKGPVCVGVDASHPSFFLYRSGVYYDPACTQKVNHGVLVIGYGDLNGKEYWLVKNSW 299

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              | ||||:|||||||||||||||:||||||
Db        300 GSNFGEQGYIRMARNKGNHCGIASYPSYPEI 330
```

RESULT 15
ABM83324
ID ABM83324 standard; protein; 313 AA.
XX
AC ABM83324;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human diagnostic and therapeutic pprotein SEQ ID NO:3573.
XX
KW gene therapy; human diagnostic and therapeutic polynucleotide; dithp.
XX
OS Homo sapiens.
XX
PN WO2004023973-A2.
XX
PD 25-MAR-2004.
XX
PF 12-SEP-2003; 2003WO-US028227.
XX
PR 12-SEP-2002; 2002US-0410259P.
PR 12-SEP-2002; 2002US-0410260P.
XX
PA (INCY-) INCYTE CORP.
XX
PI Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
PI Harthshorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV;
PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstin EH;
PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES;
PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Patury S, Shi X, Suarez CJ;
XX
DR WPI; 2004-329368/30.
DR N-PSDB; ACN41976.
XX
PT New diagnostic and therapeutic polynucleotides and polypeptides, useful
PT in diagnosing a condition, disease or disorder associated with human
PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy,or
PT in gene mapping.
XX
PS Claim 27; Page; 190pp; English.
XX
CC The invention relates to novel diagnostic and therapeutic polynucleotides
CC selected from one of the 2722 sequences defined in the specification. A
CC polynucleotide of the invention may have a use in gene therapy. The human
CC diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
CC used to diagnose a particular condition, disease or disorder associated
CC with human molecules, e.g. cell proliferative disorders,
CC autoimmune/inflammatory disorder, developmental disorder, endocrine
CC disorder, neurological disorders, gastrointestinal disorders, or
CC infections caused by virus, bacteria, fungi or parasite. The dithp
CC molecules may also be used in genetic mapping, in identifying individuals
CC from minute biological samples, in detecting single nucleotide
CC polymorphisms, as molecular weight markers, and for somatic or germline
CC gene therapy. The present sequence represents a dithp protein of the
CC invention. Note: The sequence data for this patent is not represented in
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
XX
SQ Sequence 313 AA;

Query Match 92.9%; Score 1678; DB 8; Length 313;
Best Local Similarity 94.3%; Pred. No. 1.7e-158;
Matches 312; Conservative 0; Mismatches 1; Indels 18; Gaps 1;

```
Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCST-------- 172

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
                          |||||||||||||||||| ||||||||||||||||||||||||||||||
Db        173 ----------YIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREDVLKE 222

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        223 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 282

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              ||||||||||||||||||||||||||||||
Db        283 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 313
```

Search completed: January 10, 2006, 09:39:57
Job time : 138 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 10, 2006, 09:32:19 ; Search time 39 Seconds
(without alignments)
816.609 Million cell updates/sec

Title: US-10-646-470-1
Perfect score: 1806
Sequence: 1 MKRLVCVLLVCSSAVAQLHK.....MARNKGNHCGIASFPSYPEI 331

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1791	99.2	331	2	A42482	cathepsin S (EC 3.
2	1299.5	72.0	330	2	A45087	cathepsin S (EC 3.
3	1035	57.3	217	2	S15844	cathepsin S (EC 3.
4	952.5	52.7	329	2	A49868	cathepsin K (EC 3.
5	951.5	52.7	329	2	JC2476	cathepsin K (EC 3.
6	927.5	51.4	329	2	S74227	cathepsin K (EC 3.
7	865	47.9	334	1	KHMSL	cathepsin L (EC 3.
8	859.5	47.6	334	1	KHRTL	cathepsin L (EC 3.
9	840.5	46.5	333	1	KHHUL	cathepsin L (EC 3.
10	839	46.5	334	2	A58195	cathepsin L (EC 3.
11	787	43.6	323	2	S19650	cysteine proteinas
12	785.5	43.5	326	2	S53027	cathepsin L (EC 3.
13	785	43.5	313	2	S47433	cathepsin L (EC 3.
14	780.5	43.2	339	2	A53810	cathepsin L (EC 3.
15	776.5	43.0	337	2	T24387	probable cysteine
16	768.5	42.6	344	2	JX0366	cysteine endopepti
17	762	42.2	320	2	S19651	cysteine proteinas
18	761.5	42.2	338	2	JC5443	cathepsin L-like c
19	746.5	41.3	333	1	I52525	testin precursor -
20	738	40.9	343	2	JC7183	cathepsin Q (EC 3.
21	732.5	40.6	324	2	S47432	cathepsin L (EC 3.
22	728.5	40.3	322	2	S19649	cysteine proteinas
23	724.5	40.1	331	2	JC5441	cathepsin L-like c
24	723	40.0	326	2	S43991	cathepsin L-like p
25	721.5	40.0	331	2	JC5442	cathepsin L-like c
26	696.5	38.6	218	1	KHCHL	cathepsin L (EC 3.
27	689	38.2	480	2	T01207	cysteine proteinas
28	683	37.8	326	2	T09259	cathepsin L-like p
29	679	37.6	458	1	KHRZOA	oryzain (EC 3.4.22
30	669.5	37.1	218	2	S67481	cathepsin L-like c
31	652	36.1	462	2	JN0719	drought-inducible
32	643.5	35.6	455	2	T12041	cysteine proteinas
33	643	35.6	317	2	S44151	cathepsin L (EC 3.
34	639	35.4	427	2	S57776	cysteine proteinas
35	630.5	34.9	331	2	D86413	cysteine proteinas
36	629.5	34.9	368	2	S47312	cysteine proteinas
37	625.5	34.6	374	2	T03941	cysteine proteinas
38	625	34.6	367	2	T06529	cysteine proteinas
39	621.5	34.4	454	2	JC4848	cysteine proteinas
40	619.5	34.3	464	2	S24602	cysteine proteinas
41	613.5	34.0	471	1	KHRZOB	oryzain (EC 3.4.22
42	610.5	33.8	466	2	T06416	cysteine proteinas
43	610	33.8	355	2	T05390	probable cysteine
44	607	33.6	380	1	TAGB	actinidain (EC 3.4
45	606.5	33.6	343	2	D86198	cysteine proteinas

ALIGNMENTS

RESULT 1
A42482
cathepsin S (EC 3.4.22.27) precursor - human
C;Species: Homo sapiens (man)
C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C;Accession: A42482; A53625; A42896
R;Shi, G.P.; Munger, J.S.; Meara, J.P.; Rich, D.H.; Chapman, H.A.
J. Biol. Chem. 267, 7258-7262, 1992
A;Title: Molecular cloning and expression of human alveolar macrophage cathepsin S, an e
A;Reference number: A42482; MUID:92218373; PMID:1373132
A;Accession: A42482
A;Molecule type: mRNA
A;Residues: 1-331 <SHI>
A;Cross-references: UNIPROT:P25774; UNIPARC:UPI0000151182; GB:S93414; NID:g248405; PIDN
A;Experimental source: alveolar macrophage
A;Note: sequence extracted from NCBI backbone (NCBIN:93414, NCBIP:93443)
R;Shi, G.P.; Webb, A.C.; Foster, K.E.; Knoll, J.H.M.; Lemere, C.A.; Munger, J.S.; Chapma
J. Biol. Chem. 269, 11530-11536, 1994
A;Title: Human cathepsin S: chromosomal localization, gene structure, and tissue distrib
A;Reference number: A53625; MUID:94209337; PMID:8157683
A;Accession: A53625
A;Molecule type: DNA
A;Residues: 1-210,'H',212-331 <SH2>
A;Cross-references: UNIPARC:UPI0000175CF0; GB:U07374
R;Wiederanders, B.; Bromme, D.; Kirschke, H.; von Figura, K.; Schmidt, B.; Peters, C.
J. Biol. Chem. 267, 13708-13713, 1992
A;Title: Phylogenetic conservation of cysteine proteinases. Cloning and expression of a
A;Reference number: A42896; MUID:92317106; PMID:1377692
A;Accession: A42896
A;Status: not compared with conceptual translation
A;Molecule type: mRNA; protein
A;Residues: 1-91,'M',93-160,'S',162-331 <WIE>
A;Cross-references: UNIPARC:UPI000013DFE1; GB:S39127; GB:M90696; NID:g250802; PIDN:AAA09
A;Experimental source: testis
A;Note: sequence extracted from NCBI backbone (NCBIP:107806)
C;Genetics:
A;Gene: GDB:CTSS
A;Cross-references: GDB:132414; OMIM:116845
A;Map position: 1q21-1q21
C;Superfamily: papain
C;Keywords: cysteine proteinase; hydrolase; lysosome
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-114/Domain: propeptide #status predicted <PRO>
F;115-331/Product: cathepsin S #status predicted <MAT>
F;139,278,298/Active site: Cys, His, Asn #status predicted

Query Match 99.2%; Score 1791; DB 2; Length 331;
Best Local Similarity 99.1%; Pred. No. 6.2e-142;
Matches 328; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLTSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              |||||||||||||||||||||||||||||||||||||||||:||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVTLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              |||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331


RESULT 2
A45087
cathepsin S (EC 3.4.22.27) - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 09-Jul-2004
C;Accession: A45087
R;Petanceska, S.; Devi, L.
J. Biol. Chem. 267, 26038-26043, 1992
A;Title: Sequence analysis, tissue distribution, and expression of rat cathepsin S.
A;Reference number: A45087; MUID:93100327; PMID:1281481
A;Accession: A45087
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-330 <PET>
A;Cross-references: UNIPROT:Q02765; UNIPARC:UPI00001270D0; GB:L03201; NID:g203649; PIDN:
A;Experimental source: brain
A;Note: sequence extracted from NCBI backbone (NCBIN:120879, NCBIP:120880)
C;Superfamily: papain
C;Keywords: cysteine proteinase; glycoprotein; heterodimer; hydrolase; lysosome
F;137,277,297/Active site: Cys, His, Asn #status predicted

  Query Match             72.0%;  Score 1299.5;  DB 2;  Length 330;
  Best Local Similarity   76.9%;  Pred. No. 7e-101;
  Matches  240;  Conservative   28;  Mismatches   41;  Indels    3;  Gaps    3;

Qy         22 PTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVMLHNLEHSMGMHSYDLGMNHLG 81
              ||||||| |||||  ::  ::||| ||||||||||||:|||||||||||||| :||||:|
Db         20 PTLDHHWDLWKKTRMRRNTDQNEEDVRRLIWEKNLKFIMLHNLEHSMGMHSYSVGMNHMG 79

Qy         82 DMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVDWREKGCVTEVKYQGSCGACWA 141
              ||| |||:  | |||:|  | |: | ||: |: |||||||||||||| |||||||:|||
Db         80 DMTPEEVIGYMGSLRIPRPWNRSGTLKSSSNQTLPDSVDWREKGCVTNVKYQGSCGSCWA 139

Qy        142 FSAVGALEAQLKLKTGKLVSLSAQNLVDCST-EKYGNKGCNGGFMTTAFQYIIDNKGIDS 200
              ||| |||| |||||||||||||||||||||| |||||||| ||||| |||||||   |||
Db        140 FSAEGALEGQLKLKTGKLVSLSAQNLVDCSTEEKYGNKGCGGGFMTEAFQYIIDT-SIDS 198

Qy        201 DASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKEAVANKGPVSVGV-DARHPSF 259
              :||||||||| || || | ||||||:| |||:| |: |||||| |||||||: || | ||
Db        199 EASYPYKAMDEKCLYDPKNRAATCSRYIELPFGDEEALKEAVATKGPVSVGIDDASHSSF 258

Qy        260 FLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSWGHNFGEEGYIRMARNKGNH 319
              |||:|||| :||||:|:|||||||||| |:||:|||||||| :||::|||||||  ||
Db        259 FLYQSGVYDDPSCTENMNHGVLVVGYGTLDGKDYWLVKNSWGLHFGDQGYIRMARNNKNH 318

Qy        320 CGIASFPSYPEI 331
              |||||: |||||
Db        319 CGIASYCSYPEI 330


RESULT 3
S15844
cathepsin S (EC 3.4.22.27) - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 09-Jul-2004
C;Accession: S15844; S23680; S16972; S23957
R;Ritonja, A.; Colic, A.; Dolenc, I.; Ogrinc, T.; Podobnik, M.; Turk, V.
FEBS Lett. 283, 329-331, 1991
A;Title: The complete amino acid sequence of bovine cathepsin S and a partial sequence o
A;Reference number: S15844; MUID:91257334; PMID:2044774
A;Accession: S15844
A;Molecule type: protein
A;Residues: 1-217 <RIT>
A;Cross-references: UNIPROT:P25326; UNIPARC:UPI00001270CE
R;Wiederanders, B.; Broemme, D.; Kirschke, H.; Kalkkinen, N.; Rinne, A.; Paquette, T.; T
FEBS Lett. 286, 189-192, 1991
A;Title: Primary structure of bovine cathepsin S. Comparison to cathepsins L, H, B and p
A;Reference number: S16972; MUID:91323515; PMID:1864368
A;Accession: S23680
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 22-217 <WIE>
A;Cross-references: UNIPARC:UPI000016C2D4; GB:M95211; NID:g162814; PIDN:AAA30435.1; PID:
A;Note: 143-Pro was also found
A;Accession: S16972
A;Molecule type: protein
A;Residues: 1-28;48-71;94-104;107-131;146-162;178-217 <WI2>
A;Cross-references: UNIPARC:UPI0000175CE7; UNIPARC:UPI0000175CE8; UNIPARC:UPI0000175CE9;
C;Superfamily: papain
C;Keywords: cysteine proteinase; hydrolase
F;12-110,22-66,56-99,158-206/Disulfide bonds: #status predicted
F;25,164,184/Active site: Cys, His, Asn #status predicted

  Query Match             57.3%;  Score 1035;  DB 2;  Length 217;
  Best Local Similarity   84.8%;  Pred. No. 5e-79;
  Matches  184;  Conservative   20;  Mismatches   13;  Indels    0;  Gaps    0;

Qy        115 LPDSVDWREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEK 174
              ||||:||||||||||||:||:|||||||||||||:|||||||||||||||||||| |
Db          1 LPDSMDWREKGCVTEVKYQGACGSCWAFSAVGALEAQVKLKTGKLVSLSAQNLVDCSTAK 60

Qy        175 YGNKGCNGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGR 234
              |||||||||||| |||||||| ||||:|||||||| ||||| | |||||||:| |||:|
Db         61 YGNKGCNGGFMTEAFQYIIDNNGIDSEASYPYKAMDGKCQYDVKNRAATCSRYIELPFGS 120

Qy        235 EDVLKEAVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYW 294
              |: |||||||||||||:|| | |||||::||||:|||||||||||||||:|:||:||
Db        121 EEALKEAVANKGPVSVGIDASHSSFFLYKTGVYYDPSCTQNVNHGVLVVGYGNLDGKDYW 180

Qy        295 LVKNSWGHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              ||||||| :||::|||||||| |||||||::|||||
Db        181 LVKNSWGLHFGDQGYIRMARNSGNHCGIANYPSYPEI 217


RESULT 4
A49868
cathepsin K (EC 3.4.22.-) precursor [similarity] - rabbit
N;Alternate names: osteoclast cysteine proteinase OC-2
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 30-Jun-1995 #sequence_revision 30-Jun-1995 #text_change 09-Jul-2004
C;Accession: A49868
R;Tezuka, K.; Tezuka, Y.; Maejima, A.; Sato, T.; Nemoto, K.; Kamioka, H.; Hakeda, Y.; Ku
J. Biol. Chem. 269, 1106-1109, 1994
A;Title: Molecular cloning of a possible cysteine proteinase predominantly expressed in
A;Reference number: A49868; MUID:94117413; PMID:8288568
A;Accession: A49868
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-329 <TEZ>
A;Cross-references: UNIPROT:P43236; UNIPARC:UPI00001270B5; GB:D14036; NID:g454186; PIDN:
C;Superfamily: papain
C;Keywords: cysteine proteinase; hydrolase
F;139,276,296/Active site: Cys, His, Asn #status predicted
```

```
  Query Match           52.7%;  Score 952.5;  DB 2;  Length 329;
  Best Local Similarity  54.6%;  Pred. No. 6.8e-72;
  Matches  184;  Conservative   45;  Mismatches   85;  Indels   23;  Gaps    5;

Qy          5 VCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVMLHNL 64
              | :| | | |   || :  ||  | |||||| |||  | :|  |||||||||| : :|||
Db          6 VLLLPVVSFA---LHPEEILDTQWELWKKTYSKQYNSKVDEISRRLIWEKNLKHISIHNL 62

Qy         65 EHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRIL--------- 115
              | |:|:|:|:| ||||||||||||:  |: |:||         :|: |  |
Db         63 EASLGVHTYELAMNHLGDMTSEEVVQKMTGLKVPPS-------RSHSNDTLYIPDWEGRT 115

Qy        116 PDSVDWREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKY 175
              |||:|:|:|| || || || ||:|||||:||||| ||| |||||::|| |||||| :| |
Db        116 PDSIDYRKKGYVTPVKNQGQCGSCWAFSSVGALEGQLKKKTGKLLNLSPQNLVDCVSENY 175

Qy        176 GNKGCNGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGRE 235
              || ||:|| ||||:  |:||||: :|||   |  | |:   :|| |  | |:| | |
Db        176 ---GCGGGYMTNAFQYVQRNRGIDSEDAYPYVGQDESCMYNPTGKAAKCRGYREIPEGNE 232

Qy        236 DVLKEAVANKGPVSVGVDARHPSFFLYRSGVYYEPSC-TQNVNHGVLVVGYGDLNGKEYW 294
              || |||  ||||| :||   ||  |  ||||: :| : |||| || ||||   | ::|
Db        233 KALKRAVARVGPVSVAIDASLTSFQFYSKGVYYDENCSSDNVNHAVLAVGYGIQKGNKHW 292

Qy        295 LVKNSWGHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              ::||||| ::| :||| ||||| | ||||:  |:|::
Db        293 IIKNSWGESWGNKGYILMARNKNNACGIANLASFPKM 329
```

RESULT 5
JC2476
cathepsin K (EC 3.4.22.-) precursor - human
N;Alternate names: cathepsin O2
C;Species: Homo sapiens (man)
C;Date: 21-Feb-1995 #sequence_revision 05-Apr-1995 #text_change 09-Jul-2004
C;Accession: JC2476; S55763; S68459; I38752; S48830
R;Inaoka, T.; Bilbe, G.; Ishibashi, O.; Tezuka, K.; Kumegawa, M.; Kokubo, T.
Biochem. Biophys. Res. Commun. 206, 89-96, 1995
A;Title: Molecular cloning of human cDNA for cathepsin K: Novel cysteine proteinase pred
A;Reference number: JC2476; MUID:95118380; PMID:7818555
A;Accession: JC2476
A;Molecule type: mRNA
A;Residues: 1-329 <INA>
A;Cross-references: UNIPROT:P43235; UNIPARC:UPI0000000DFB; EMBL:X82153; NID:g562756; PID
R;Broemme, D.; Okamoto, K.
Biol. Chem. Hoppe-Seyler 376, 379-384, 1995
A;Title: Human cathepsin O2, a novel cysteine protease highly expressed in osteoclastoma
A;Reference number: S55763; MUID:96082523; PMID:7576232
A;Accession: S55763
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-329 <BRO>
A;Cross-references: UNIPARC:UPI0000000DFB; GB:S79895; NID:g1195555; PIDN:AAB35521.1; PID
R;Shi, G.P.; Chapman, H.A.; Bhairi, S.M.; DeLeeuw, C.; Reddy, V.Y.; Weiss, S.J.
FEBS Lett. 357, 129-134, 1995
A;Title: Molecular cloning of human cathepsin O, a novel endoproteinase and homologue of
A;Reference number: I38752; MUID:95104457; PMID:7805878
A;Accession: S68459
A;Molecule type: mRNA
A;Residues: 1-329 <SHI>
A;Cross-references: UNIPARC:UPI0000000DFB; EMBL:U13665; NID:g606922; PIDN:AAA65233.1; PI
A;Experimental source: tissue-type blood
C;Genetics:
A;Gene: GDB:CTSK; PYCD; CTSO1; CTSO2; PKND; CTSO
A;Cross-references: GDB:453910; OMIM:265800; OMIM:600550; OMIM:601105
A;Map position: 1q21-1q21
C;Superfamily: papain
C;Keywords: cysteine proteinase; glycoprotein; hydrolase
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-114/Domain: propeptide #status predicted <PRO>
F;115-329/Product: cathepsin K #status predicted <MAT>
F;103,161,213/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;139,276,296/Active site: Cys, His, Asn #status predicted

```
  Query Match           52.7%;  Score 951.5;  DB 2;  Length 329;
  Best Local Similarity  54.9%;  Pred. No. 8.3e-72;
  Matches  184;  Conservative   46;  Mismatches   86;  Indels   19;  Gaps    5;

Qy          5 VCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVMLHNL 64
              | :| | | |   |: :  || || |||||: |||  | :|  ||||||||||:: :|||
Db          6 VLLLPVVSFA---LYPEEILDTHWELWKKTHRKQYNNKVDEISRRLIWEKNLKYISIHNL 62

Qy         65 EHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRIL-------PD 117
              | |:|:|:|:| ||||||||||||:  |: |:||    |     ||    :       ||
Db         63 EASLGVHTYELAMNHLGDMTSEEVVQKMTGLKVPLSHSR-----SNDTLYIPEWEGRAPD 117

Qy        118 SVDWREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGN 177
              |||:|:|| || || || ||:|||||:||||| ||| |||||::|| |||||| :|   |
Db        118 SVDYRKKGYVTPVKNQGQCGSCWAFSSVGALEGQLKKKTGKLLNLSPQNLVDCVSE---N 174

Qy        178 KGCNGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDV 237
              || ||:|| ||||:  |:||||: :|||   :  | |:   :|| |  | |:| | |
Db        175 DGCGGGYMTNAFQYVQKNRGIDSEDAYPYVGQEESCMYNPTGKAAKCRGYREIPEGNEKA 234

Qy        238 LKEAVANKGPVSVGVDARHPSFFLYRSGVYYEPSC-TQNVNHGVLVVGYGDLNGKEYWLV 296
              || |||  ||||| :||   ||  |  ||||: || : |:|| || ||||   | ::|::
Db        235 LKRAVARVGPVSVAIDASLTSFQFYSKGVYYDESCNSDNLNHAVLAVGYGIQKGNKHWII 294

Qy        297 KNSWGHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              ||||| |:| :||| ||||| | ||||:  |:|::
Db        295 KNSWGENWGNKGYILMARNKNNACGIANLASFPKM 329
```

RESULT 6
S74227
cathepsin K (EC 3.4.22.-) precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 29-Jan-1998 #sequence_revision 13-Feb-1998 #text_change 09-Jul-2004
C;Accession: S74227
R;Rantakokko, J.; Aro, H.T.; Savontaus, M.; Vuorio, E.
FEBS Lett. 393, 307-313, 1996
A;Title: Mouse cathepsin K: cDNA cloning and predominant expression of the gene in osteo
A;Reference number: S74227; MUID:96409328; PMID:8814310
A;Accession: S74227
A;Molecule type: mRNA
A;Residues: 1-329 <RAN>
A;Cross-references: UNIPROT:P55097; UNIPARC:UPI0000044DB0; EMBL:X94444; NID:g1149524; PI
A;Experimental source: strain c57/bl; tissue type calvaria
C;Superfamily: papain
C;Keywords: cysteine proteinase; glycoprotein; hydrolase
F;1-17/Domain: signal sequence #status predicted <SIG>
F;18-114/Domain: propeptide #status predicted <PRO>
F;115-329/Product: cathepsin K #status predicted <MAT>
F;103,213/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;136-177,170-210,269-318/Disulfide bonds: #status predicted
F;139,276,296/Active site: Cys, His, Asn #status predicted

```
  Query Match           51.4%;  Score 927.5;  DB 2;  Length 329;
  Best Local Similarity  53.7%;  Pred. No. 8.4e-70;
  Matches  176;  Conservative   49;  Mismatches   96;  Indels    7;  Gaps    4;

Qy          7 VLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVMLHNLEH 66
              |||:   : | |  :  ||  | |||||: |||  | :|  |||||||||| :  ||||
Db          6 VLLLPMVSFA-LSPEEMLDTQWELWKKTHQKQYNSKVDEISRRLIWEKNLKQISAHNLEA 64

Qy         67 SMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVP--SQWQRNITYKSNPNRILPDSVDWREK 124
              |:|:|:|:| ||||||||||||:  |: ||:|    :  :  |        :|||:|:|:|
Db         65 SLGVHTYELAMNHLGDMTSEEVVQKMTGLRIPPSRSYSNDTLYTPEWEGRVPDSIDYRKK 124

Qy        125 GCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGCNGGF 184
              | || || || ||:|||||: |||| ||| |||||::|| |||||| || |   || ||:
Db        125 GYVTPVKNQGQCGSCWAFSSAGALEGQLKKKTGKLLALSPQNLVDCVTENY---GCGGGY 181

Qy        185 MTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKEAVAN 244
              |||||||:  | ||||: ::||   |  | |::  :|| |  | |:| | |  || |||
```

```
Db        182 MTTAFQYVQQNGGIDSEDAFPYVGQDESCMYNATAKAAKCRGYREIPVGNEKALKRAVAR 241

Qy        245 KGPVSVGVDARHPSFFLYRSGVYYEPSCTQ-NVNHGVLVVGYGDLNGKEYWLVKNSWGHN 303
              ||:|| :||   ||  |  ||||: :| : |||| |||||||   | ::|::||||| :
Db        242 VGPISVSIDASLASFQFYSRGVYYDENCDRDNVNHAVLVVGYGTQKGSKHWIIKNSWGES 301

Qy        304 FGEEGYIRMARNKGNHCGIASFPSYPEI 331
              :| :||  :|||| | ||| :  |:|::
Db        302 WGNKGYALLARNKNNACGITNMASFPKM 329
```

RESULT 7
KHMSL
cathepsin L (EC 3.4.22.15) precursor - mouse
N;Alternate names: major excreted protein (MEP); procathepsin L
C;Species: Mus musculus (house mouse)
C;Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 09-Jul-2004
C;Accession: S01177; A34972; A25999; A32333; A45927; S13890; S48734; S64672
R;Troen, B.R.; Gal, S.; Gottesman, M.M.
Biochem. J. 246, 731-735, 1987
A;Title: Sequence and expression of the cDNA for MEP (major excreted protein), a transfo
A;Reference number: S01177; MUID:88076849; PMID:3689328
A;Accession: S01177
A;Molecule type: mRNA
A;Residues: 1-334 <TRO1>
A;Cross-references: UNIPROT:P06797; UNIPARC:UPI0000003E31; EMBL:X06086; NID:g53046; PIDN
A;Accession: A34972
A;Molecule type: protein
A;Residues: 18-28 <TRO2>
A;Cross-references: UNIPARC:UPI0000172C60
R;Portnoy, D.A.; Erickson, A.H.; Kochan, J.; Ravetch, J.V.; Unkeless, J.C.
J. Biol. Chem. 261, 14697-14703, 1986
A;Title: Cloning and characterization of a mouse cysteine proteinase.
A;Reference number: A25999; MUID:87033683; PMID:3533924
A;Accession: A25999
A;Molecule type: mRNA
A;Residues: 1-334 <POR>
A;Cross-references: UNIPARC:UPI0000003E31; EMBL:J02583; NID:g192681; PIDN:AAA37445.1; PI
R;Joseph, L.J.; Chang, L.C.; Stamenkovich, D.; Sukhatme, V.P.
J. Clin. Invest. 81, 1621-1629, 1988
A;Title: Complete nucleotide and deduced amino acid sequences of human and murine prepro
A;Reference number: A92768; MUID:88213715; PMID:2835398
A;Accession: A32333
A;Molecule type: mRNA
A;Residues: 1-57,'I',59-334 <JOS>
A;Cross-references: UNIPARC:UPI00000E830F; GB:M20495; NID:g200500; PIDN:AAA39984.1; PID:
R;Denhardt, D.T.; Hamilton, R.T.; Parfett, C.L.J.; Edwards, D.R.; St.Pierre, R.; Waterho
Cancer Res. 46, 4590-4593, 1986
A;Title: Close relationship of the major excreted protein of transformed murine fibrobla
A;Reference number: A45927; MUID:86271744; PMID:3755373
A;Accession: A45927
A;Molecule type: mRNA
A;Residues: 89-300 <DEN>
A;Cross-references: UNIPARC:UPI000016CEE1; GB:X04392; NID:g53050; PIDN:CAA27980.1; PID:g
R;Stearns, N.A.; Dong, J.; Pan, J.X.; Brenner, D.A.; Sahagian, G.G.
Arch. Biochem. Biophys. 283, 447-457, 1990
A;Title: Comparison of cathepsin L synthesized by normal and transformed cells at the ge
A;Reference number: S13890; MUID:91112761; PMID:2275556
A;Accession: S13890
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-334 <STE>
A;Cross-references: UNIPARC:UPI0000003E31
R;Ishidoh, K.; Kominami, E.
FEBS Lett. 352, 281-284, 1994
A;Title: Multi-step processing of procathepsin L in vitro.
A;Reference number: S48734; MUID:95010724; PMID:7925987
A;Accession: S48734
A;Molecule type: protein
A;Residues: 104-124 <ISH>
A;Cross-references: UNIPARC:UPI0000172C61
R;Jean, D.; Hermann, J.; Rodrigues-Lima, F.; Barel, M.; Balbo, M.; Frade, R.
Biochem. J. 312, 961-969, 1995
A;Title: Identification on melanoma cells of p39, a cysteine proteinase that cleaves C3,
A;Reference number: S64672; MUID:96128086; PMID:8554545
A;Accession: S64672
A;Status: preliminary
A;Molecule type: protein
A;Residues: 18-34;273-292;295-313 <JEA>
A;Cross-references: UNIPARC:UPI0000172C62; UNIPARC:UPI0000172C63; UNIPARC:UPI0000172C64
C;Function:
A;Description: catalyzes hydrolysis of peptide bonds in proteins
A;Pathway: intracellular protein degradation
A;Note: important role in the lysosomal degradation of proteins
C;Superfamily: papain
C;Keywords: cysteine proteinase; glycoprotein; heterodimer; hydrolase; lysosome; protein
F;1-17/Domain: signal sequence #status predicted <SIG>
F;18-113/Domain: propeptide #status predicted <PRO>
F;114-288/Product: cathepsin L heavy chain #status predicted <HCH>
F;291-334/Product: cathepsin L light chain #status predicted <LCH>
F;135-178,169-211,269-322/Disulfide bonds: #status predicted
F;138,276,300/Active site: Cys, His, Asn #status predicted
F;221,268/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match   47.9%; Score 865; DB 1; Length 334;
  Best Local Similarity   49.1%; Pred. No. 1.4e-64;
  Matches 168; Conservative   58; Mismatches   96; Indels   20; Gaps   8;

```
Qy          1 MKRLVCVLLVC-SSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFV 59
              |  |: : ::|  :|:|    | |    || || |: : |   |||  || |||||:: :
Db          1 MNLLLLLAVLCLGTALATPKFDQTFSAEWHQWKSTHRRLY-GTNEEEWRRAIWEKNMRMI 59

Qy         60 MLHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQ-----WQRNITYKSNPNRI 114
               ||| |:| | | : : ||  ||||:||   ::: |         :|  :  |
Db         60 QLHNGEYSNGQHGFSMEMNAFGDMTNEEFRQVVNGYRHQKHKKGRLFQEPLMLK------ 113

Qy        115 LPDSVDWREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEK 174
              :| |||||||||| || || ||:|||||| | || |: |||||:||| |||||||
Db        114 IPKSVDWREKGCVTPVKNQGQCGSCWAFSASGCLEGQMFLKTGKLISLSEQNLVDCS-HA 172

Qy        175 YGNKGCNGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGR 234
               ||:||||| |  ||||| :| |:||: ||||:| |  |:| ::: |  : : ::|  :
Db        173 QGNQGCNGGLMDFAFQYIKENGGLDSEESYPYEAKDGSCKYRAEFAVANDTGFVDIPQ-Q 231

Qy        235 EDVLKEAVANKGPVSVGVDARHPSFFLYRSGVYYEPSC-TQNVNHGVLVVGYG----DLN 289
              |  | :||| ||:|| :|| |||   | ||:||||:| ::|::||||:||||    | |
Db        232 EKALMKAVATVGPISVAMDASHPSLQFYSSGIYYEPNCSSKNLDHGVLLVGYGYEGTDSN 291

Qy        290 GKEYWLVKNSWGHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
               :|||||||  :| ||||::|::: ||||:|:  ||| :
Db        292 KNKYWLVKNSWGSEWGMEGYIKIAKDRDNHCGLATAASYPVV 333
```

RESULT 8
KHRTL
cathepsin L (EC 3.4.22.15) precursor - rat
N;Alternate names: cyclic protein-2; major excreted protein (MEP); procathepsin L
C;Species: Rattus norvegicus (Norway rat)
C;Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 09-Jul-2004
C;Accession: S07098; S00155; S02445; A41550; S02446
R;Ishidoh, K.; Kominami, E.; Suzuki, K.; Katunuma, N.
FEBS Lett. 259, 71-74, 1989
A;Title: Gene structure and 5'-upstream sequence of rat cathepsin L.
A;Reference number: S07098; MUID:90092543; PMID:2599113
A;Accession: S07098
A;Molecule type: DNA
A;Residues: 1-334 <ISH1>
A;Cross-references: UNIPROT:P07154; UNIPARC:UPI0000172C5D; EMBL:X51648; NID:g57532
A;Note: only part of the nucleotide sequence is given
R;Ishidoh, K.; Towatari, T.; Imajoh, S.; Kawasaki, H.; Kominami, E.; Katunuma, N.; Suzuk
FEBS Lett. 223, 69-73, 1987
A;Title: Molecular cloning and sequencing of cDNA for rat cathepsin L.
A;Reference number: S00155; MUID:88030047; PMID:3666143
A;Accession: S00155
A;Molecule type: mRNA
A;Residues: 1-30,'Q',32-237,'P',239-334 <ISH2>

A;Cross-references: UNIPARC:UPI0000167A40; EMBL:Y00697; NID:g55887; PIDN:CAA68691.1; PID
R;Towatari, T.; Katunuma, N.
FEBS Lett. 236, 57-61, 1988
A;Title: Amino acid sequence of rat liver cathepsin L.
A;Reference number: S02445; MUID:88296890; PMID:3402618
A;Accession: S02445
A;Molecule type: protein
A;Residues: 114-288;291-334 <TOW>
A;Cross-references: UNIPARC:UPI0000172C5E; UNIPARC:UPI0000172C5F
R;Erickson-Lawrence, M.; Zabludoff, S.D.; Wright, W.W.
Mol. Endocrinol. 5, 1789-1798, 1991
A;Title: Cyclic protein-2, a secretory product of rat sertoli cells, is the proenzyme fo
A;Reference number: A41550; MUID:92168015; PMID:1791830
A;Accession: A41550
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 88-334 <ERI>
A;Cross-references: UNIPARC:UPI0000170CE0; GB:S85184; NID:g246147; PIDN:AAB21516.1; PID:
C;Genetics:
A;Introns: 42/3; 83/3; 132/3; 207/3; 262/1; 301/2
C;Complex: heterodimer of disulfide linked chains produced from a single chain precursor
C;Function:
A;Description: catalyzes hydrolysis of peptide bonds in proteins
A;Pathway: intracellular protein degradation
A;Note: important role in the lysosomal degradation of proteins
C;Superfamily: papain
C;Keywords: cysteine proteinase; glycoprotein; heterodimer; hydrolase; lysosome; protein
F;1-17/Domain: signal sequence #status predicted <SIG>
F;18-113/Domain: propeptide #status predicted <PRO>
F;114-288/Product: cathepsin L heavy chain #status experimental <HCH>
F;291-334/Product: cathepsin L light chain #status experimental <LCH>
F;135-178,169-211,269-322/Disulfide bonds: #status predicted
F;138,276,300/Active site: Cys, His, Asn #status predicted
F;221/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 47.6%; Score 859.5; DB 1; Length 334;
Best Local Similarity 48.7%; Pred. No. 4.1e-64;
Matches 165; Conservative 59; Mismatches 90; Indels 25; Gaps 7;

```
Qy          4 LVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVMLHN 63
              |:  :|   :|:|    | | :  || || |: : |   |||  || :||||:: : |||
Db          5 LLLAVLCLGTALATPKFDQTFNAQWHHWKSTHRRLY-GTNEEEWRRAVWEKNMRMIQLHN 63

Qy         64 LEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRI--------L 115
               |:| | | : : ||  ||||:||   :::  |          |     |:       :
Db         64 GEYSNGKHGFTMEMNAFGDMTNEEFRQIVNGYR---------HQKHKKGRLFQEPLMLQI 114

Qy        116 PDSVDWREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKY 175
              | :|||||||||| || || ||:|||||| | || |: ||||||:||| ||||||| ::
Db        115 PKTVDWREKGCVTPVKNQGQCGSCWAFSASGCLEGQMFLKTGKLISLSEQNLVDCSHDQ- 173

Qy        176 GNKGCNGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGRE 235
              ||:||||| |  ||||| :| |:||: ||||:| |  |:| ::|  |  : : ::|  :|
Db        174 GNQGCNGGLMDFAFQYIKENGGLDSEESYPYEAKDGSCKYRAEYAVANDTGFVDIPQ-QE 232

Qy        236 DVLKEAVANKGPVSVGVDARHPSFFLYRSGVYYEPSC-TQNVNHGVLVVGYG----DLNG 290
              | :|||  ||:|| :|| |||   | ||:||||:| :::::|||||||||    | |
Db        233 KALMKAVATVGPISVAMDASHPSLQFYSSGIYYEPNCSSKDLDHGVLVVGYGYEGTDSNK 292

Qy        291 KEYWLVKNSWGHNFGEEGYIRMARNKGNHCGIASFPSYP 329
               :|||||||||  :| :|||::|::: ||||:|:  |||
Db        293 DKYWLVKNSWGKEWGMDGYIKIAKDRNNHCGLATAASYP 331
```

RESULT 9
KHHUL
cathepsin L (EC 3.4.22.15) precursor [validated] - human
N;Alternate names: major excreted protein (MEP); procathepsin L
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 09-Jul-2004
C;Accession: S01002; B32333; S09065; A45043; S00323; B27011; A26069; A32683; D27011; F27
R;Gal, S.; Gottesman, M.M.
Biochem. J. 253, 303-306, 1988

A;Title: Isolation and sequence of a cDNA for human pro-(cathepsin L).
A;Reference number: S01002; MUID:88339905; PMID:3421948
A;Accession: S01002
A;Molecule type: mRNA
A;Residues: 1-333 <GAL>
A;Cross-references: UNIPROT:P07711; UNIPARC:UPI000004CAA2; GB:X12451; NID:g29714; PIDN:(
R;Joseph, L.J.; Chang, L.C.; Stamenkovich, D.; Sukhatme, V.P.
J. Clin. Invest. 81, 1621-1629, 1988
A;Title: Complete nucleotide and deduced amino acid sequences of human and murine prepro
A;Reference number: A92768; MUID:88213715; PMID:2835398
A;Accession: B32333
A;Molecule type: mRNA
A;Residues: 1-333 <JOS>
A;Cross-references: UNIPARC:UPI000004CAA2; GB:M20496; NID:g809235; PIDN:AAA66974.1; PID:
R;Joseph, L.; Lapid, S.; Sukhatme, V.
Nucleic Acids Res. 15, 3186, 1987
A;Title: The major ras induced protein in NIH3T3 cells is cathepsin L.
A;Reference number: S09065; MUID:87174843; PMID:3550705
A;Accession: S09065
A;Molecule type: mRNA
A;Residues: 113-154 <JO2>
A;Cross-references: UNIPARC:UPI000016A66C; EMBL:X05256; NID:g29718; PIDN:CAA28877.1; PID
R;Chauhan, S.S.; Popescu, N.C.; Ray, D.; Fleischmann, R.; Gottesman, M.M.; Troen, B.R.
J. Biol. Chem. 268, 1039-1045, 1993
A;Title: Cloning, genomic organization, and chromosomal localization of human cathepsin
A;Reference number: A45043; MUID:93123212; PMID:8419312
A;Accession: A45043
A;Molecule type: DNA
A;Residues: 40-46;82-86;130-135;205-210;259-264;299-304 <CHA>
A;Cross-references: UNIPARC:UPI0000172C53; UNIPARC:UPI0000172C54; UNIPARC:UPI0000172C55,
A;Note: only exon-intron splice junctions are shown
R;Ritonja, A.; Popovic, T.; Kotnik, M.; Machleidt, W.; Turk, V.
FEBS Lett. 228, 341-345, 1988
A;Title: Amino acid sequences of the human kidney cathepsins H and L.
A;Reference number: S00322; MUID:88137635; PMID:3342889
A;Accession: S00323
A;Molecule type: protein
A;Residues: 114-147,'P',148-220,'X',222-267,'N',269-288;292-333 <RIT>
A;Cross-references: UNIPARC:UPI0000111D17; UNIPARC:UPI0000172C59
R;Machleidt, W.; Ritonja, A.; Popovic, T.; Kotnik, M.; Brzin, J.; Turk, V.; Machleidt, I
in Cysteine Proteinases and Their Inhibitors, Turk, V., ed., pp.3-18, Walter de Gruyter,
A;Title: Human cathepsins B, H and L: characterization by amino acid sequences and some
A;Reference number: A27011
A;Accession: B27011
A;Molecule type: protein
A;Residues: 'X',115-129,'M',131-133,'E',135-141;292-307,'TD',310-333 <MA2>
A;Cross-references: UNIPARC:UPI0000172C5A; UNIPARC:UPI0000172C5B
R;Mason, R.W.; Walker, J.E.; Northrop, F.D.
Biochem. J. 240, 373-377, 1986
A;Title: The N-terminal amino acid sequences of the heavy and light chains of human cath
A;Reference number: A26069; MUID:87127952; PMID:3545185
A;Accession: A26069
A;Molecule type: protein
A;Residues: 114-147,'P',149-152,'Y';292-333 <MAS>
A;Cross-references: UNIPARC:UPI0000111D17; UNIPARC:UPI0000172C5C
R;Smith, S.M.; Gottesman, M.M.
J. Biol. Chem. 264, 20487-20495, 1989
A;Title: Activity and deletion analysis of recombinant human cathepsin L expressed in Es
A;Reference number: A32683; MUID:90062183; PMID:2684978
A;Contents: annotation
C;Genetics:
A;Gene: GDB:CTSL
A;Cross-references: GDB:119824; OMIM:116880
A;Map position: 9q22.1-9q22.2
A;Introns: 42/3; 83/3; 132/3; 207/3; 262/1; 301/2
C;Complex: heterodimer of disulfide linked chains produced from a single chain precursor
C;Function:
A;Description: catalyzes hydrolysis of peptide bonds in proteins
A;Pathway: intracellular protein degradation
A;Note: important role in the lysosomal degradation of proteins
C;Superfamily: papain
C;Keywords: cysteine proteinase; glycoprotein; heterodimer; hydrolase; lysosome; protein
F;1-17/Domain: signal sequence #status predicted <SIG>

```
F;18-113/Domain: propeptide #status predicted <PRO>
F;114-333/Product: cathepsin L #status experimental <MAT>
F;114-288/Product: cathepsin L heavy chain #status experimental <HCH>
F;292-333/Product: cathepsin L light chain #status experimental <LCH>
F;135-178,169-211,269-322/Disulfide bonds: #status predicted
F;138,276,300/Active site: Cys, His, Asn #status predicted
F;221/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match            46.5%;  Score 840.5;  DB 1;  Length 333;
  Best Local Similarity  48.5%;  Pred. No. 1.6e-62;
  Matches  164;  Conservative   57;  Mismatches   98;  Indels   19;  Gaps    8;

Qy          4 LVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVMLHN 63
              |:          | | |  | :|:  |  ||  : : |   |||  || :||||:| : |||
Db          5 LILAAFCLGIASATLTFDHSLEAQWTKWKAMHNRLY-GMNEEGWRRAVWEKNMKMIELHN 63

Qy         64 LEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSL--RVPSQ---WQRNITYKSNPNRILPDS 118
                |:  | ||: : ||  |||||||   :|:    | | :   :|  : |::      | |
Db         64 QEYREGKHSFTMAMNAFGDMTSEEFRQVMNGFQNRKPRKGKVFQEPLFYEA------PRS 117

Qy        119 VDWREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNK 178
              ||||||| || || || ||:|||||| |||| |:  |||:|:||| |||||||  : ||:
Db        118 VDWREKGYVTPVKNQGQCGSCWAFSATGALEGQMFRKTGRLISLSEQNLVDCSGPQ-GNE 176

Qy        179 GCNGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVL 238
              ||||| |  |||| : || |:||: ||||:| :  |:|: ||  |  : : ::|  :|  |
Db        177 GCNGGLMDYAFQYVQDNGGLDSEESYPYEATEESCKYNPKYSVANDTGFVDIP-KQEKAL 235

Qy        239 KEAVANKGPVSVGVDARHPSFFLYRSGVYYEPSC-TQNVNHGVLVVGYG----DLNGKEY 293
              :|||  ||:|| :|| | ||  |: |:|:|| | :::::|||||||||    : :  :|
Db        236 MKAVATVGPISVAIDAGHESFLFYKEGIYFEPDCSSEDMDHGVLVVGYGFESTESDNNKY 295

Qy        294 WLVKNSWGHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              ||||||||  :|  ||::||::: |||||||  ||| :
Db        296 WLVKNSWGEEWGMGGYVKMAKDRRNHCGIASAASYPTV 333


RESULT 10
A58195
cathepsin L (EC 3.4.22.15) precursor - pig
N;Alternate names: major excreted protein (MEP); procathepsin L
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 16-Oct-1996 #sequence_revision 18-Oct-1996 #text_change 09-Jul-2004
C;Accession: A58195; S59911; S59916
R;Okamura, N.; Tamba, M.; Uchiyama, Y.; Sugita, Y.; Dacheux, F.; Syntin, P.; Dacheux, J.
Biochim. Biophys. Acta 1245, 221-226, 1995
A;Title: Direct evidence for the elevated synthesis and secretion of procathepsin L in t
A;Reference number: S59911; MUID:96069841; PMID:7492581
A;Accession: A58195
A;Molecule type: mRNA
A;Residues: 1-334 <OKA>
A;Cross-references: UNIPROT:Q28944; UNIPARC:UPI00001270BB; DDBJ:D37917; NID:g710655; PID
A;Experimental source: epididymis
A;Note: the authors translated the codon CCG for residue 203 as Thr and CCC for residue
A;Note: submitted to the DDBJ/EMBL/GenBank databases by Naomichi Okamura, 8 August 1994
A;Accession: S59911
A;Molecule type: protein
A;Residues: 18-31 <OK2>
A;Cross-references: UNIPARC:UPI0000175CF1
C;Complex: heterodimer of disulfide linked chains produced from a single chain precursor
C;Function:
A;Description: catalyzes hydrolysis of peptide bonds in proteins
A;Pathway: protein degradation
A;Note: important role in the lysosomal degradation of proteins
C;Superfamily: papain
C;Keywords: cysteine proteinase; glycoprotein; heterodimer; hydrolase; lysosome; protein
F;1-17/Domain: signal sequence #status predicted <SIG>
F;18-113/Domain: propeptide #status predicted <PRO>
F;114-334/Product: cathepsin L #status predicted <MAT>
F;135-178,169-212,270-323/Disulfide bonds: #status predicted
F;138,277,301/Active site: Cys, His, Asn #status predicted
F;222,292/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match            46.5%;  Score 839;  DB 2;  Length 334;
  Best Local Similarity  48.8%;  Pred. No. 2.1e-62;
  Matches  163;  Conservative   58;  Mismatches  101;  Indels   12;  Gaps    8;

Qy          4 LVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVMLHN 63
              |  : |  :||  :|  |  ||  |: || |:|: |   |||  || :||||:| : |||
Db          7 LTALCLGIASAAPKL--DQNLDADWYKWKATHGRLY-GMNEEGWRRAVWEKNMKMIELHN 63

Qy         64 LEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVDWRE 123
               |:| | | : :  ||  ||||:||   :|:  :      :  : ::|    : | ||||||
Db         64 QEYSQGKHGFSMAMNAFGDMTNEEFRQVMNGFQNQKHKKGKVFHESLVLEV-PKSVDWRE 122

Qy        124 KGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGCNGG 183
              || || || || ||:|||||| |||| |:  |||||||| |||||||  : ||:|||||
Db        123 KGYVTAVKNQGQCGSCWAFSATGALEGQMFRKTGKLVSLSEQNLVDCSRPQ-GNQGCNGG 181

Qy        184 FMTTAFQYIIDNKGIDSDASYPYKAMDL-KCQYDSKYRAATCSKYTELPYGREDVLKEAV 242
              |  ||||: || |:|:: ||||   :   | |  :  ||  : : ::|  ||  | :||
Db        182 LMDNAFQYVKDNGGLDTEESYPYLGRETNSCTYKPECSAANDTGFVDIPQ-REKALMKAV 240

Qy        243 ANKGPVSVGVDARHPSFFLYRSGVYYEPSC-TQNVNHGVLVVGYG----DLNGKEYWLVK 297
              |  ||:|| :|| | ||  |:||:||:| | :::::|||||||||    | |  ::|:||
Db        241 ATVGPISVAIDAGHSSFQFYKSGIYYDPDCSSKDLDHGVLVVGYGFEGTDSNSSKFWIVK 300

Qy        298 NSWGHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              ||||  :|  ||::||::: |||||::  ||| :
Db        301 NSWGPEWGWNGYVKMAKDQNNHCGISTAASYPTV 334


RESULT 11
S19650
cysteine proteinase (EC 3.4.22.-) precursor (clone LCP2) - American lobster
C;Species: Homarus americanus (American lobster)
C;Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 09-Jul-2004
C;Accession: S19650; S31655
R;Laycock, M.V.; MacKay, R.M.; Di Fruscio, M.; Gallant, J.W.
FEBS Lett. 292, 115-120, 1991
A;Title: Molecular cloning of three cDNAs that encode cysteine proteinases in the digest
A;Reference number: S19649; MUID:92070467; PMID:1959590
A;Accession: S19650
A;Molecule type: mRNA
A;Residues: 1-323 <LAY>
A;Cross-references: UNIPROT:P25782; UNIPARC:UPI0000128CAC; EMBL:X63568; NID:g11052; PIDN
R;Laycock, M.V.; MacKay, R.M.; Di Fruscio, M.; Gallant, J.W.
FEBS Lett. 301, 125, 1992
A;Title: Correction. Molecular cloning of three cDNAs that encode cysteine proteinases i
A;Reference number: S31654; MUID:93083613; PMID:1451782
A;Accession: S31655
A;Molecule type: mRNA
A;Residues: 1-323 <LA2>
A;Cross-references: UNIPARC:UPI0000128CAC; EMBL:X63568; NID:g11052; PIDN:CAA45128.1; PID
C;Superfamily: papain
C;Keywords: cysteine proteinase; disulfide bond; hydrolase; zymogen
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-107/Domain: propeptide #status predicted <PRO>
F;108-323/Product: cysteine proteinase #status predicted <MAT>
F;128-171,162-204,263-312/Disulfide bonds: #status predicted
F;131,270,290/Active site: Cys, His, Asn #status predicted

  Query Match            43.6%;  Score 787;  DB 2;  Length 323;
  Best Local Similarity  46.6%;  Pred. No. 4.5e-58;
  Matches  156;  Conservative   55;  Mismatches   98;  Indels   26;  Gaps    5;

Qy          5 VCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVMLHNL 64
              | || :|  |:|           |  :|  ||:|| :  |:: ||:|:|:| |::   |
Db          3 VAVLFLCGVALA------AASPSWEHFKGKYGRQYVDAEEDSYRRVIFEQNQKYIEEFNK 56

Qy         65 EHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRIL--------- 115
              ::  |  :::| ||  |||| ||  ::|           ||  :| |  :
Db         57 KYENGEVTFNLAMNKFGDMTLEEFNAVMKG---------NIPRRSAPVSVFYPKKETGPQ 107

Qy        116 PDSVDWREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKY 175
                |||| || || || || ||:|||||  |:|| |  |||| |:||: | |||||   |
```

```
Db        108 ATEVDWRTKGAVTPVKDQGQCGSCWAFSTTGSLEGQHFLKTGSLISLAEQQLVDCS-RPY 166

Qy        176 GNKGCNGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGRE 235

Db        167 GPQGCNGGWMNDAFDYIKANNGIDTEAAYPYEARDGSCRFDSNSVAATCSGHTNIASGSE 226

Qy        236 DVLKEAVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQN-VNHGVLVVGYGDLNGKEYW 294

Db        227 TGLQQAVRDIGPISVTIDAAHSSFQFYSSGVYYEPSCSPSYLDHAVLAVGYGSEGGQDFW 286

Qy        295 LVKNSWGHNFGEEGYIRMARNKGNHCGIASFPSYP 329

Db        287 LVKNSWATSWGDAGYIKMSRNRNNNCGIATVASYP 321


RESULT 12
S53027
cathepsin L (EC 3.4.22.15) precursor - penaeid shrimp (Penaeus vannamei) (fragment)
C;Species: Penaeus vannamei
C;Date: 08-Jul-1995 #sequence_revision 03-Nov-1995 #text_change 09-Jul-2004
C;Accession: S53027
R;le Boulay, C.; van Wormhoudt, A.; Sellos, D.
submitted to the EMBL Data Library, March 1995
A;Description: Cloning and sequencing of crustacean cathepsin-L-like cysteine proteases
A;Reference number: S53027
A;Accession: S53027
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-326 <LEB>
A;Cross-references: UNIPROT:Q27760; UNIPARC:UPI00000790FA; EMBL:X85127; NID:g728636; PID
C;Superfamily: papain
C;Keywords: cysteine proteinase; hydrolase
F;133,272,293/Active site: Cys, His, Asn #status predicted

  Query Match             43.5%;  Score 785.5;  DB 2;  Length 326;
  Best Local Similarity   48.3%;  Pred. No. 6e-58;
  Matches  159;  Conservative   56;  Mismatches  101;  Indels   13;  Gaps    7;

Qy          4 LVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVMLHN 63

Db          6 LACVVAV---AVA----TPSLRQQWQNFKAEHGRRYASVQEERYRLSVFEQNQQFIDDHN 58

Qy         64 LEHSMGMHSYDLGMNHLGDMTSEEVMSLMSS-LRVPSQWQRNITYKSNPNRILPDSVDWR 122

Db         59 ARFENGEVTFTLQMNQFGDMTSEEIVATMNGFLGAPT--RRPAAVLKADDETLPEKVDWR 116

Qy        123 EKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGCNG 182

Db        117 TKGAVTPVKDQKQCGSCWAFSTTGSLEGQHFLKDGKLVSLSEQNLVDCS-DKFGNMGCMG 175

Qy        183 GFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKEAV 242

Db        176 GLMDQAFRYIKANKGIDTEDSYPYEAQDGKCRFDASNVGATDTGYVDVEHGSESALKKAV 235

Qy        243 ANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQN-VNHGVLVVGYG-DLNGKEYWLVKNSW 300

Db        236 ATIGPISVGIDASQSTFHFYHTGVYHDDHCSSTMLDHGVLAVGYGSDENGGDFWLVKNSW 295

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYP 329

Db        296 NTSWGDKGYIKMSRNRNNNCGIASQASYP 324


RESULT 13
S47433
cathepsin L (EC 3.4.22.15) - Norway lobster
C;Species: Nephrops norvegicus (Norway lobster)
C;Date: 23-Nov-1994 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
C;Accession: S47433
R;le Boulay, C.; van Wormhoudt, A.; Sellos, D.
submitted to the EMBL Data Library, August 1994
A;Description: Molecular cloning and sequencing of the two cDNAs that encode cathepsin L
A;Reference number: S47432
A;Accession: S47433
A;Molecule type: mRNA
A;Residues: 1-313 <LEB>
A;Cross-references: UNIPROT:Q27708; UNIPARC:UPI0000084019; EMBL:X80990; NID:g530735; PID
C;Superfamily: papain
C;Keywords: cysteine proteinase; hydrolase
F;122,260,280/Active site: Cys, His, Asn #status predicted

  Query Match             43.5%;  Score 785;  DB 2;  Length 313;
  Best Local Similarity   47.5%;  Pred. No. 6.3e-58;
  Matches  153;  Conservative   58;  Mismatches  101;  Indels   10;  Gaps    5;

Qy         11 CSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVMLHNLEHSMGM 70

Db          1 CGLALA------TASPSWEHFKTQYGRKYGDAKEELYRQRVFQQNEQLVEAFNKKFENGE 54

Qy         71 HSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVDWREKGCVTEV 130

Db         55 VTFKVAMNQFGDMTNEEFNAVMKGYKKGSRGEPTTVFTAE-GRPMAADVDWRTKGAVTPV 113

Qy        131 KYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGCNGGFMTTAFQ 190

Db        114 KDQGQCGSCWAFSATGSLEGQHFLKNNELVSLSEQELVDCSTE-YGNDGCGGGWMTSAFD 172

Qy        191 YIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKEAVANKGPVSV 250

Db        173 YIKDNGGIDTESSYPYEAQDRSCRFDANSIGATCTGFVEVQH-TEEALHEAVSDIGPISV 231

Qy        251 GVDARHPSFFLYRSGVYYEPSCT-QNVNHGVLVVGYGDLNGKEYWLVKNSWGHNFGEEGY 309

Db        232 AIDASHFSFQFYSSGVYYEKKCSPTNLDHGVLAVGYGTESTEDYWLVKNSWGSGWGDAGY 291

Qy        310 IRMARNKGNHCGIASFPSYPEI 331

Db        292 IKMSRNRDNNCGIASEPSYPTV 313


RESULT 14
A53810
cathepsin L (EC 3.4.22.15) precursor - flesh fly (Sarcophaga peregrina)
C;Species: Sarcophaga peregrina
C;Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 09-Jul-2004
C;Accession: A53810
R;Homma, K.; Kurata, S.; Natori, S.
J. Biol. Chem. 269, 15258-15264, 1994
A;Title: Purification, characterization, and cDNA cloning of procathepsin L from the cul
he differentiation of imaginal discs.
A;Reference number: A53810; MUID:94253090; PMID:8195162
A;Accession: A53810
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-339 <HOM>
A;Cross-references: UNIPROT:Q26636; UNIPARC:UPI00001270BD; GB:D16533; NID:g505139; PIDN
C;Superfamily: papain
C;Keywords: cysteine proteinase; hydrolase
F;146,285,306/Active site: Cys, His, Asn #status predicted

  Query Match             43.2%;  Score 780.5;  DB 2;  Length 339;
  Best Local Similarity   48.5%;  Pred. No. 1.7e-57;
  Matches  166;  Conservative   43;  Mismatches  116;  Indels   17;  Gaps    7;

Qy          3 RLVCVLLVCSSAVAQ-LHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVML 61

Db          2 RTVLVALLALVALTQAISPLDLIKEEWHTYKLQHRKNYANEVEERFRMKIFNENRHKIAK 61

Qy         62 HNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMS----SLRVPSQWQR------NITYKSNP 111

Db         62 HNQLFAQGKVSYKLGLNKYADMLHHEFKETMNGYNHTLR---QLMRERTGLVGATYIPPA 118

Qy        112 NRILPDSVDWREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCS 171

Db        119 HVTVPKSVDWREHGAVTGVKDQGHCGSCWAFSSTGALEGQHFRKAGVLVSLSEQNLVDCS 178
```

```
Qy        172 TEKYGNKGCNGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELP 231
              | |||| ||||| |  ||:|| || |||:: ||||: :|  | ::      || : : ::|
Db        179 T-KYGNNGCNGGLMDNAFRYIKDNGGIDTEKSYPYEGIDDSCHFNKATIGATDTGFVDIP 237

Qy        232 YGREDVLKEAVANKGPVSVGVDARHPSFFLYRSGVYYEPSC-TQNVNHGVLVVGYG-DLN 289
               | |: :|:|||  ||||| :|| | || ||  ||| || |  ||::|||||||| | :
Db        238 EGDEEKMKKAVATMGPVSVAIDASHESFQLYSEGVYNEPECDEQNLDHGVLVVGYGTDES 297

Qy        290 GKEYWLVKNSWGHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              | :|||||||||  :||:|||:||||: | ||||:  ||| :
Db        298 GMDYWLVKNSWGTTWGEQGYIKMARNQNNQCGIATASSYPTV 339


RESULT 15
T24387
probable cysteine proteinase (EC 3.4.22.-) T03E6.7 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T24387
R;Lloyd, C.
submitted to the EMBL Data Library, March 1997
A;Reference number: Z19885
A;Accession: T24387
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-337 <WIL>
A;Cross-references: UNIPROT:O45734; UNIPARC:UPI0000077623; EMBL:Z92812; PIDN:CAB07275.1;
A;Experimental source: clone T03E6
C;Genetics:
A;Gene: CESP:T03E6.7
A;Map position: 5
A;Introns: 86/3; 119/3; 269/1
C;Superfamily: papain
C;Keywords: cysteine proteinase; hydrolase
F;144,283,304/Active site: Cys, His, Asn #status predicted

  Query Match             43.0%;  Score 776.5;  DB 2;  Length 337;
  Best Local Similarity   45.5%;  Pred. No. 3.6e-57;
  Matches  150;  Conservative   64;  Mismatches  109;  Indels    7;  Gaps    6;

Qy          4 LVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVMLHN 63
              ||  ::  :||       :  ::  |  :|: : |:| |  |:      : ||:  :  ||
Db          9 LVAAVVAVNSAKLSRQIESAIE-KWDDYKEDFDKEYSESEEQTYME-AFVKNMIHIENHN 66

Qy         64 LEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRV--PSQWQRNITYKSNPNRILPDSVDW 121
               :| :|  ::::|:||: |:   :   |    |:   |: : : :: :  |  :|| |||
Db         67 RDHRLGRKTFEMGLNHIADLPFSQYRKLNGYRRLFGDSRIKNSSSFLAPFNVQVPDEVDW 126

Qy        122 REKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGCN 181
              |:   ||:|| || ||:|||||| |||| |   | |:||||| |||||||| |||| |||
Db        127 RDTHLVTDVKNQGMCGSCWAFSATGALEGQHARKLGQLVSLSEQNLVDCST-KYGNHGCN 185

Qy        182 GGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKEA 241
              || |  ||:|| || |:|:: |||||  |:|| :: |   |    | : | | |: || |
Db        186 GGLMDQAFEYIRDNHGVDTEESYPYKGRDMKCHFNKKTVGADDKGYVDTPEGDEEQLKIA 245

Qy        242 VANKGPVSVGVDARHPSFFLYRSGVYYEPSC-TQNVNHGVLVVGYG-DLNGKEYWLVKNS 299
              || :||:|: :|| | || ||: ||||:  | :: ::||||:|||| |    :||:||||
Db        246 VATQGPISIAIDAGHRSFQLYKKGVYYDEECSSEELDHGVLLVGYGTDPEHGDYWIVKNS 305

Qy        300 WGHNFGEEGYIRMARNKGNHCGIASFPSYP 329
              ||  :||:||||:|||: ||||:|:  |||
Db        306 WGAGWGEKGYIRIARNRNNHCGVATKASYP 335


Search completed: January 10, 2006, 09:36:32
Job time : 40 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 10, 2006, 09:32:19 ; Search time 45 Seconds
(without alignments)
608.126 Million cell updates/sec

Title: US-10-646-470-1
Perfect score: 1806
Sequence: 1 MKRLVCVLLVCSSAVAQLHK.....MARNKGNHCGIASFPSYPEI 331

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/6_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/H_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/RE_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1806	100.0	331	1	US-08-208-007A-8	Sequence 8, Appli
2	1806	100.0	331	2	US-09-701-685-2	Sequence 2, Appli
3	1806	100.0	331	2	US-09-290-586A-20	Sequence 20, Appl
4	1791	99.2	331	2	US-08-860-255A-2	Sequence 2, Appli
5	1791	99.2	331	2	US-08-915-095A-8	Sequence 8, Appli
6	1791	99.2	331	2	US-08-798-096-8	Sequence 8, Appli
7	1791	99.2	331	2	US-08-798-095A-8	Sequence 8, Appli
8	1791	99.2	331	2	US-09-953-956-8	Sequence 8, Appli
9	1791	99.2	331	2	US-08-553-125A-8	Sequence 8, Appli
10	1791	99.2	331	2	US-10-114-464-8	Sequence 8, Appli
11	1784	98.8	331	1	US-08-330-121B-4	Sequence 4, Appli
12	1784	98.8	331	2	US-08-536-861-4	Sequence 4, Appli
13	1784	98.8	331	4	PCT-US95-13820-4	Sequence 4, Appli
14	1700.5	94.2	330	2	US-10-010-577-2	Sequence 2, Appli
15	1599	88.5	331	2	US-10-010-580-2	Sequence 2, Appli
16	953.5	52.8	329	1	US-08-806-959-2	Sequence 2, Appli
17	952.5	52.7	329	1	US-08-208-007A-7	Sequence 7, Appli
18	952.5	52.7	329	1	US-08-330-121B-3	Sequence 3, Appli
19	952.5	52.7	329	2	US-08-915-095A-7	Sequence 7, Appli
20	952.5	52.7	329	2	US-08-798-096-7	Sequence 7, Appli
21	952.5	52.7	329	2	US-08-798-095A-7	Sequence 7, Appli
22	952.5	52.7	329	2	US-09-953-956-7	Sequence 7, Appli
23	952.5	52.7	329	2	US-08-553-125A-7	Sequence 7, Appli
24	952.5	52.7	329	2	US-08-536-861-3	Sequence 3, Appli
25	952.5	52.7	329	2	US-10-114-464-7	Sequence 7, Appli
26	952.5	52.7	329	4	PCT-US95-13820-3	Sequence 3, Appli
27	951.5	52.7	329	1	US-08-330-121B-2	Sequence 2, Appli
28	951.5	52.7	329	1	US-08-852-807-20	Sequence 20, Appl
29	951.5	52.7	329	2	US-08-964-308-4	Sequence 4, Appli
30	951.5	52.7	329	2	US-08-964-313-4	Sequence 4, Appli
31	951.5	52.7	329	2	US-08-860-255A-1	Sequence 1, Appli
32	951.5	52.7	329	2	US-09-069-138-4	Sequence 4, Appli
33	951.5	52.7	329	2	US-08-915-095A-2	Sequence 2, Appli
34	951.5	52.7	329	2	US-08-798-096-2	Sequence 2, Appli
35	951.5	52.7	329	2	US-08-798-095A-2	Sequence 2, Appli
36	951.5	52.7	329	2	US-09-953-956-2	Sequence 2, Appli
37	951.5	52.7	329	2	US-08-553-125A-2	Sequence 2, Appli
38	951.5	52.7	329	2	US-08-536-861-2	Sequence 2, Appli
39	951.5	52.7	329	2	US-10-114-464-2	Sequence 2, Appli
40	951.5	52.7	329	2	US-09-290-586A-19	Sequence 19, Appl
41	951.5	52.7	329	4	PCT-US95-13820-2	Sequence 2, Appli
42	945.5	52.4	329	1	US-08-208-007A-2	Sequence 2, Appli
43	945.5	52.4	329	4	PCT-US94-04781-2	Sequence 2, Appli
44	944.5	52.3	329	2	US-08-684-932A-36	Sequence 36, Appl
45	942.5	52.2	329	2	US-08-964-308-15	Sequence 15, Appl

ALIGNMENTS

```
RESULT 1
US-08-208-007A-8
; Sequence 8, Application US/08208007A
; Patent No. 5501969
;  GENERAL INFORMATION:
;    APPLICANT:  HASTINGS, ET AL.
;    TITLE OF INVENTION:  Human Osteoclast-Derived Cathepsin
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;      ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;      STREET:  6 BECKER FARM ROAD
;      CITY:  ROSELAND
;      STATE:  NEW JERSEY
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH DISKETTE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/208,007A
;      FILING DATE:  March 8, 1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  No. 5501969e
;      FILING DATE:  No. 5501969e
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FERRARO, GREGORY D.
;      REGISTRATION NUMBER:  36,134
;      REFERENCE/DOCKET NUMBER:  325800-95
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  331 AMINO ACIDS
;      TYPE:  AMINO ACID
;      STRANDEDNESS:
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  PROTEIN
US-08-208-007A-8

  Query Match             100.0%;  Score 1806;  DB 1;  Length 331;
  Best Local Similarity   100.0%;  Pred. No. 4e-176;
  Matches  331;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              |||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331


RESULT 2
US-09-701-685-2
; Sequence 2, Application US/09701685
; Patent No. 6387629
; GENERAL INFORMATION:
;  APPLICANT: Schneider, Patrick
;  APPLICANT:  Yamamoto, Karen K.
;  APPLICANT:  French, Cynthia K.
;  APPLICANT:  Reprogen, Inc.
;  TITLE OF INVENTION: Use of Cathepsin S in the Diagnosis and Treatment of
;  TITLE OF INVENTION:  Endometriosis
;  FILE REFERENCE: 018002-001310US
;  CURRENT APPLICATION NUMBER: US/09/701,685
;  CURRENT FILING DATE:  2001-03-26
;  PRIOR APPLICATION NUMBER: WO PCT/US99/12335
;  PRIOR FILING DATE: 1999-06-03
;  PRIOR APPLICATION NUMBER: US 60/088,017
;  PRIOR FILING DATE: 1998-06-04
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 331
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-701-685-2

  Query Match            100.0%;  Score 1806;  DB 2;  Length 331;
  Best Local Similarity  100.0%;  Pred. No. 4e-176;
  Matches  331;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              |||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331


RESULT 3
US-09-290-586A-20
; Sequence 20, Application US/09290586A
; Patent No. 6800473
; GENERAL INFORMATION:
;  APPLICANT: SANTAMARIA, Ignio
;  APPLICANT:  VELASCO, Gloria
;  APPLICANT:  CAZORLA, Maite
;  APPLICANT:  FUEYO, Antonio
;  APPLICANT:  CAMPO, Elias
;  APPLICANT:  LOPEZ-OTIN, Carlos
;  APPLICANT:  AOKI, Takanori
;  APPLICANT:  IWATA, Kazushi
;  TITLE OF INVENTION: NOVEL HUMAN CATHEPSIN L2 PROTEIN, GENE ENCODING SAID
;  TITLE OF INVENTION:  PROTEIN AND USE THEREOF
;  FILE REFERENCE: 99-410A/WMC/01332
;  CURRENT APPLICATION NUMBER: US/09/290,586A
;  CURRENT FILING DATE:  1999-04-13
;  PRIOR APPLICATION NUMBER: JP 10-172147
;  PRIOR FILING DATE: 1998-06-05
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
;   LENGTH: 331
;   TYPE: PRT
;   ORGANISM: Human
US-09-290-586A-20

  Query Match            100.0%;  Score 1806;  DB 2;  Length 331;
  Best Local Similarity  100.0%;  Pred. No. 4e-176;
  Matches  331;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              |||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331


RESULT 4
US-08-860-255A-2
; Sequence 2, Application US/08860255A
; Patent No. 6274336
; GENERAL INFORMATION:
;  APPLICANT:  Abdel-Mequid, Sherin
;  APPLICANT:  Desjarlais, Renee
;  APPLICANT:  Janson, Cheryl
;  APPLICANT:  Smith, Ward
;  APPLICANT:  Zhao, Baoguang
;  TITLE OF INVENTION:  Method of Inhibiting Cathepsin K
```

```
;  FILE REFERENCE:  P50574-X1
;  CURRENT APPLICATION NUMBER: US/08/860,255A
;  CURRENT FILING DATE:  1997-06-26
;  PRIOR APPLICATION NUMBER:  60/008,108
;  PRIOR FILING DATE:  1995-10-30
;  PRIOR APPLICATION NUMBER:  60/007,473
;  PRIOR FILING DATE:  1995-11-22
;  PRIOR APPLICATION NUMBER:  60/008,992
;  PRIOR FILING DATE:  1995-12-21
;  PRIOR APPLICATION NUMBER:  60/013,748
;  PRIOR FILING DATE:  1996-03-20
;  PRIOR APPLICATION NUMBER:  60/013,764
;  PRIOR FILING DATE:  1996-03-20
;  PRIOR APPLICATION NUMBER:  60/013,747
;  PRIOR FILING DATE:  1996-03-20
;  PRIOR APPLICATION NUMBER:  60/017,455
;  PRIOR FILING DATE:  1996-05-17
;  PRIOR APPLICATION NUMBER:  60/017,892
;  PRIOR FILING DATE:  1996-05-17
;  PRIOR APPLICATION NUMBER:  60/020,478
;  PRIOR FILING DATE:  1996-06-13
;  PRIOR APPLICATION NUMBER:  60/022,047
;  PRIOR FILING DATE:  1996-07-22
;  PRIOR APPLICATION NUMBER:  60/023,494
;  PRIOR FILING DATE:  1996-08-07
;  PRIOR APPLICATION NUMBER:  60/023,742
;  PRIOR FILING DATE:  1996-08-08
;  NUMBER OF SEQ ID NOS:  7
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 331
;   TYPE: PRT
;   ORGANISM: homo sapiens
US-08-860-255A-2

  Query Match             99.2%;  Score 1791;  DB 2;  Length 331;
  Best Local Similarity   99.1%;  Pred. No. 1.4e-174;
  Matches  328;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLTSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||:||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVTLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              |||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331


RESULT 5
US-08-915-095A-8
; Sequence 8, Application US/08915095A
; Patent No. 6383793
; GENERAL INFORMATION:
;  APPLICANT:  Hastings, et al.
;  TITLE OF INVENTION:  HUMAN OSTEOCLAST-DERIVED CATHEPSIN
;  FILE REFERENCE:  PF107D4
;  CURRENT APPLICATION NUMBER: US/08/915,095A
;  CURRENT FILING DATE:  1997-08-20
;  NUMBER OF SEQ ID NOS:  14
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 8
;   LENGTH: 331
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-08-915-095A-8

  Query Match             99.2%;  Score 1791;  DB 2;  Length 331;
  Best Local Similarity   99.1%;  Pred. No. 1.4e-174;
  Matches  328;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLTSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||:||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVTLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              |||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331


RESULT 6
US-08-798-096-8
; Sequence 8, Application US/08798096
; Patent No. 6387682
; GENERAL INFORMATION:
;  APPLICANT:  Hastings, et al.
;  TITLE OF INVENTION:  HUMAN OSTEOCLAST-DERIVED CATHEPSIN
;  FILE REFERENCE:  PF107D2
;  CURRENT APPLICATION NUMBER: US/08/798,096
;  CURRENT FILING DATE:  1997-02-12
;  NUMBER OF SEQ ID NOS:  14
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 8
;   LENGTH: 331
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-08-798-096-8

  Query Match             99.2%;  Score 1791;  DB 2;  Length 331;
  Best Local Similarity   99.1%;  Pred. No. 1.4e-174;
  Matches  328;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLTSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||:||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVTLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
```

```
                ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
                |||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331


RESULT 7
US-08-798-095A-8
; Sequence 8, Application US/08798095A
; Patent No. 6423507
; GENERAL INFORMATION:
;  APPLICANT:  Hastings, et al.
;  TITLE OF INVENTION:  HUMAN OSTEOCLAST-DERIVED CATHEPSIN
;  FILE REFERENCE:  PF107D3
;  CURRENT APPLICATION NUMBER: US/08/798,095A
;  CURRENT FILING DATE:  1997-02-12
;  NUMBER OF SEQ ID NOS:  14
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 8
;    LENGTH: 331
;    TYPE: PRT
;    ORGANISM: Homo sapiens
US-08-798-095A-8

  Query Match             99.2%;  Score 1791;  DB 2;  Length 331;
  Best Local Similarity   99.1%;  Pred. No. 1.4e-174;
  Matches  328;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
                |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLTSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
                |||||||||||||||||||||||||||||||||||||:||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVTLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
                |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
                |||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331


RESULT 8
US-09-953-956-8
; Sequence 8, Application US/09953956
; Patent No. 6475487
; GENERAL INFORMATION:
;  APPLICANT: Hastings, et al.
;  TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
;  FILE REFERENCE: PF107D2D1
;  CURRENT APPLICATION NUMBER: US/09/953,956
;  CURRENT FILING DATE:  2001-09-18
;  PRIOR APPLICATION NUMBER: 09/219,441
;  PRIOR FILING DATE: 1998-12-23
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
;    LENGTH: 331
;    TYPE: PRT
;    ORGANISM: Homo sapiens
US-09-953-956-8

  Query Match             99.2%;  Score 1791;  DB 2;  Length 331;
  Best Local Similarity   99.1%;  Pred. No. 1.4e-174;
  Matches  328;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
                |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLTSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
                |||||||||||||||||||||||||||||||||||||:||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVTLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
                |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
                |||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331


RESULT 9
US-08-553-125A-8
; Sequence 8, Application US/08553125A
; Patent No. 6475766
; GENERAL INFORMATION:
;  APPLICANT:  Hastings, et al.
;  TITLE OF INVENTION:  HUMAN OSTEOCLAST-DERIVED CATHEPSIN
;  FILE REFERENCE:  PF107D1
;  CURRENT APPLICATION NUMBER: US/08/553,125A
;  CURRENT FILING DATE:  1995-11-07
;  NUMBER OF SEQ ID NOS:  14
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 8
;    LENGTH: 331
;    TYPE: PRT
;    ORGANISM: Homo sapiens
US-08-553-125A-8

  Query Match             99.2%;  Score 1791;  DB 2;  Length 331;
  Best Local Similarity   99.1%;  Pred. No. 1.4e-174;
  Matches  328;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
                |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLTSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
                |||||||||||||||||||||||||||||||||||||:||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVTLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
                |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREDVLKE 240
```

```
Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              |||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331


RESULT 10
US-10-114-464-8
; Sequence 8, Application US/10114464
; Patent No. 6680375
; GENERAL INFORMATION:
;  APPLICANT: Hastings, et al.
;  TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
;  FILE REFERENCE: PF107D5
;  CURRENT APPLICATION NUMBER: US/10/114,464
;  CURRENT FILING DATE:  2002-04-03
;  PRIOR APPLICATION NUMBER: 08/553,125
;  PRIOR FILING DATE: 1995-11-07
;  PRIOR APPLICATION NUMBER: 08/208,007
;  PRIOR FILING DATE: 1994-03-08
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
;   LENGTH: 331
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-114-464-8

  Query Match             99.2%;  Score 1791;  DB 2;  Length 331;
  Best Local Similarity   99.1%;  Pred. No. 1.4e-174;
  Matches  328;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLTSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||:||||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVTLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREDVLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              |||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331


RESULT 11
US-08-330-121B-4
; Sequence 4, Application US/08330121B
; Patent No. 5736357
;  GENERAL INFORMATION:
;    APPLICANT:  Bromme, Dieter
;    APPLICANT:  Okamoto, Kathleen
;    TITLE OF INVENTION:  CATHEPSIN O2 PROTEASE
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton
;      ADDRESSEE:  & Herbert
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/330,121B
;      FILING DATE:  27-OCT-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silva, Robin M.
;      REGISTRATION NUMBER:  38,304
;      REFERENCE/DOCKET NUMBER:  A-60261/DJB/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  331 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-330-121B-4

  Query Match             98.8%;  Score 1784;  DB 1;  Length 331;
  Best Local Similarity   99.1%;  Pred. No. 7.2e-174;
  Matches  328;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||  |||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREVDLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              |||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331


RESULT 12
US-08-536-861-4
; Sequence 4, Application US/08536861
; Patent No. 6544767
;  GENERAL INFORMATION:
;    APPLICANT:  Bromme, Dieter
;    APPLICANT:  Okamoto, Kathleen
;    TITLE OF INVENTION:  CATHEPSIN O2 PROTEASE
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
```

```
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/536,861
;      FILING DATE:  02-OCT-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silva, Robin M.
;      REGISTRATION NUMBER:  38,304
;      REFERENCE/DOCKET NUMBER:  A-60261-1/DJB/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  331 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-536-861-4

  Query Match             98.8%;  Score 1784;  DB 2;  Length 331;
  Best Local Similarity   99.1%;  Pred. No. 7.2e-174;
  Matches  328;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              |||||||||||||||||||||||||||| |||||||||||||||||||||||  |||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREVDLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              |||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331


RESULT 13
PCT-US95-13820-4
; Sequence 4, Application PC/TUS9513820
;  GENERAL INFORMATION:
;    APPLICANT:  Khepri Pharmaceuticals, Inc.
;    TITLE OF INVENTION:  CATHEPSIN O2 PROTEASE
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/13820
;      FILING DATE:  26-OCT-1995
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US UNKNOWN
;      FILING DATE:  02-OCT-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/330,121
;      FILING DATE:  27-OCT-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silva, Robin M.
;      REGISTRATION NUMBER:  38,304
;      REFERENCE/DOCKET NUMBER:  FP-60261-1-PC/DJB/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  331 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
PCT-US95-13820-4

  Query Match             98.8%;  Score 1784;  DB 4;  Length 331;
  Best Local Similarity   99.1%;  Pred. No. 7.2e-174;
  Matches  328;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              |||||||||||||||||||||||||||| |||||||||||||||||||||||  |||||
Db        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDQKCQYDSKYRAATCSKYTELPYGREVDLKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              |||||||||||||||||||||||||||||||
Db        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331


RESULT 14
US-10-010-577-2
; Sequence 2, Application US/10010577
; Patent No. 6784288
; GENERAL INFORMATION:
;  APPLICANT: thurmond, robin l
;  APPLICANT:  baker, sherry
;  APPLICANT:  karlsson, lars
;  TITLE OF INVENTION: polynucleotide and polypeptide sequences of monkey cathepsin s
;  FILE REFERENCE: ORT1457
;  CURRENT APPLICATION NUMBER: US/10/010,577
;  CURRENT FILING DATE:  2001-11-08
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn version 3.1
```

```
; SEQ ID NO 2
;   LENGTH: 330
;   TYPE: PRT
;   ORGANISM: monkey
US-10-010-577-2

  Query Match             94.2%;  Score 1700.5;  DB 2;  Length 330;
  Best Local Similarity   93.7%;  Pred. No. 2.5e-165;
  Matches  310;  Conservative   11;  Mismatches    9;  Indels    1;  Gaps    1;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              ||:|||| |||||| |||||||||||:|||||||||||||||||||||||||||||||
Db          1 MKQLVCVLFVCSSAVTQLHKDPTLDHHWNLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              ||||||||||||||||||||||||||||||||||||:|||||||||||||::||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPNQWQRNITYKSNPNQMLPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCS-EKYGNKGC 179

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              |||||| |||||||||||||:||||||| | |||||||||||||||||||||||||||
Db        180 NGGFMTEAFQYIIDNKGIDSEASYPYKATDQKCQYDSKYRAATCSKYTELPYGREDVLKE 239

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              |||||||| ||||| |||||||||||||:|:||| |||||||:|||||||||||||||
Db        240 AVANKGPVCVGVDASHPSFFLYRSGVYYDPACTQKVNHGVLVIGYGDLNGKEYWLVKNSW 299

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              | ||||:||||||||||||||:||||||
Db        300 GSNFGEQGYIRMARNKGNHCGIASYPSYPEI 330


RESULT 15
US-10-010-580-2
; Sequence 2, Application US/10010580
; Patent No. 6780985
; GENERAL INFORMATION:
;  APPLICANT: thurmond, robin l
;  APPLICANT:  baker, sherry
;  APPLICANT:  karlsson, lars
;  TITLE OF INVENTION: polynucleotide and polypeptide sequences of canine cathepsin s
;  FILE REFERENCE: ORT-1458
;  CURRENT APPLICATION NUMBER: US/10/010,580
;  CURRENT FILING DATE:  2001-11-08
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
;   LENGTH: 331
;   TYPE: PRT
;   ORGANISM: canine
US-10-010-580-2

  Query Match             88.5%;  Score 1599;  DB 2;  Length 331;
  Best Local Similarity   87.9%;  Pred. No. 6e-155;
  Matches  291;  Conservative   19;  Mismatches   21;  Indels    0;  Gaps    0;

Qy          1 MKRLVCVLLVCSSAVAQLHKDPTLDHHWHLWKKTYGKQYKEKNEEAVRRLIWEKNLKFVM 60
              || || :| :|| ||||:|||||||||:|||||| |||||:|||  ||||||||||||
Db          1 MKWLVGLLPLCSYAVAQVHKDPTLDHHWNLWKKTYSKQYKEENEEVARRLIWEKNLKFVM 60

Qy         61 LHNLEHSMGMHSYDLGMNHLGDMTSEEVMSLMSSLRVPSQWQRNITYKSNPNRILPDSVD 120
              |||||||||||||||||||||| |||:||| |||||||||||:||:|| |: ||||||
Db         61 LHNLEHSMGMHSYDLGMNHLGDMTGEEVISLMGSLRVPSQWQRNVTYRSNSNQKLPDSVD 120

Qy        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 WREKGCVTEVKYQGSCGACWAFSAVGALEAQLKLKTGKLVSLSAQNLVDCSTEKYGNKGC 180

Qy        181 NGGFMTTAFQYIIDNKGIDSDASYPYKAMDLKCQYDSKYRAATCSKYTELPYGREDVLKE 240
              |||||||||||||| ||||:||||||||: ||:|||| |||||||||||:| || |||
Db        181 NGGFMTTAFQYIIDNNGIDSEASYPYKAMNGKCRYDSKKRAATCSKYTELPFGSEDALKE 240

Qy        241 AVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQNVNHGVLVVGYGDLNGKEYWLVKNSW 300
              |||||||||| :|| | |||||||||||||||||||||||||||||:||||:||||||||
Db        241 AVANKGPVSVAIDASHYSFFLYRSGVYYEPSCTQNVNHGVLVVGYGNLNGKDYWLVKNSW 300

Qy        301 GHNFGEEGYIRMARNKGNHCGIASFPSYPEI 331
              | |||::|||||||| |||||||||:||||||
Db        301 GLNFGDQGYIRMARNSGNHCGIASYPSYPEI 331


Search completed: January 10, 2006, 09:37:29
Job time : 46 secs
```